Spring 5-14-2019

Systematic Comparison of Parameter Estimation Approaches Using the Generalized-growth Model for Prediction of Epidemic Outbreaks

Yiseul Lee

Kimberlyn M. Roosa

Amna Tariq

Gerardo Chowell

ABSTRACT

Systematic Comparison of Parameter Estimation Approaches Using the Generalized-growth Model for Prediction of Epidemic Outbreaks

By

Yiseul Lee, Kimberlyn Roosa, Amna Tariq, and Gerardo Chowell

May 7th, 2019

Background- Many different mathematical models are used to assess and predict the outbreaks. The model is selected by the characteristics of the outbreaks. Here, we utilize the generalized growth model (GGM), one of the simplest mathematical models, with the real outbreaks to compare two parameter estimation methods.
Materials and Methods- 25 outbreaks are used to analyze. We use GGM with the ascending phase of each outbreak and estimated the r and p parameters with both the least square (LSQ) and maximum likelihood estimation (MLE) methods. For both parameter estimation methods, we conduct the parametric bootstrap method to construct the confidence interval of parameters. We compare the two estimation methods by the RMSE, Anscombe residual, and prediction coverage.
Results- The result show that the most outbreaks have similar r and p parameters, RMSE, Anscombe, and prediction coverage for LSQ and MLE. Although Anscombe values for LSQ are higher than the values for MLE, the different between results of the two methods are minimal for the most outbreaks.
Conclusion- The study is shown that LSQ and MLE do not result in different values of the parameter estimation, RMSE, Anscombe, and prediction coverage with GGM.

SCHOOL BOARD PERCEPTIONS OF RESPONSIBILITIES FOR CHILDHOOD OVERWEIGHT

by

JOAN Q. STUDENT

B.A., GEORGIA STATE UNIVERSITY

(List other degrees awarded in the same format)

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

APPROVAL PAGE

SCHOOL BOARD PERCEPTIONS OF RESPONSIBILITIES FOR CHILDHOOD OVERWEIGHT

by

JOAN Q. STUDENT

Approved:

__Dr. Gerardo Chowell____
Committee Chair

__Dr. Ruiyan Luo_____
Committee Member

__May 7, 2019_____
Date

Acknowledgments

(This page is optional.)
Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Yiseul Lee
Signature of Author

TABLE OF CONTENTS

List of Tables

Table 1 Result of r and p parameters with 95% CI, RMSE, Anscombe residual, prediction coverage, and the length of ascending phase by LSQ for each outbreak.

LSQ	r (95% CI)	p (95% CI)	RMSE	Anscombe	Prediction interval coverage (%)	length of ascending phase
Zika (Antioquia, 2016)	0.97 (0.69, 1.3)	0.57 (0.5, 0.65)	4.6547	26.1701	96.67	30 /104 days
Zika (Antioquia, 2016)	1.6 (0.71, 2.9)	0.43 (0.24, 0.67)	3.7148	16.3014	100.00	15/104 days
Zika (Antioquia, 2016)	1.4 (0.78, 2.3)	0.47 (0.31, 0.65)	3.3072	16.4055	100.00	16/104 days
FMD (UK,2001-120days)	0.56 (0.37, 0.78)	0.7 (0.6, 0.82)	3.2558	38.0777	96.00	25/229 days
Ebola (Tonkolili, 2014)	0.19 (0.079, 0.42)	0.77 (0.51, 1)	4.7434	8.2564	83.33	6/69 weeks
Cholera (Aalborg,1853)	0.55 (0.36, 0.8)	0.78 (0.69, 0.88)	7.8581	36.7622	90.00	20/108 days
Ebola (Bo, 2014)	0.13 (0.083, 0.2)	0.81 (0.69, 0.94)	9.6592	27.4445	80.00	10/67 weeks
Ebola (Bombali, 2014)	0.081 (0.064, 0.13)	0.95 (0.81, 1)	4.0774	17.0611	87.50	8/64 weeks
Ebola (Bomi, 2014)	1.2 (0.46, 2)	0.13 (0, 0.38)	5.9372	19.6657	75.00	8/66 weeks
Ebola (Congo, 1976)	1.3 (0.65, 2.1)	0.44 (0.28, 0.63)	3.6671	19.5928	100.00	20/52 days
Ebola (Grand Bassa, 2014)	0.42 (0.15, 0.97)	0.34 (0.058, 0.63)	3.5119	7.6974	100.00	9/64 weeks
Ebola (Gueckedou, 2014)	0.14 (0.04, 0.32)	0.65 (0.37, 1)	3.8494	18.0131	90.91	11/90 weeks
Ebola (Kenema, 2014)	0.57 (0.33, 0.85)	0.47 (0.34, 0.6)	9.1378	17.5478	87.50	8/70 weeks
Ebola (Margibi, 2014)	0.2 (0.14, 0.26)	0.75 (0.67, 0.84)	15.7639	67.8336	50.00	10/68 weeks
Ebola (Margibi, 2014)	0.21 (0.16, 0.28)	0.72 (0.65, 0.8)	17.9671	73.2763	54.55	11/68 weeks
Ebola (Montserrado, 2014)	0.088 (0.081, 0.11)	0.98 (0.9, 1)	9.3648	47.311	50.00	10/71 weeks
Ebola (Port Loko, 2014)	0.57 (0.36, 0.83)	0.5 (0.4, 0.61)	8.8882	2.8635	100.00	8/64 weeks
Ebola (Uganda, 2000)	0.34 (0.21, 0.53)	0.67 (0.52, 0.82)	5.1478	1.9795	100.00	6/18 weeks
Ebola (Western Area Rural, 2014)	0.32 (0.22, 0.44)	0.62 (0.53, 0.71)	7.064	12.5064	90.00	10/63 weeks
Ebola (Western Area Urban, 2014)	0.5 (0.32, 0.73)	0.53 (0.44, 0.63)	7.1063	12.1428	90.00	10/62 weeks
FMD (Uruguay, 2001)	9.2 (7.5, 11)	0.52 (0.49, 0.55)	78.1816	163.4178	36.36	11/27 days
HIV-AIDS (Japan, 1985-2012)	0.1 (0.097, 0.11)	0.5 (0.49, 0.51)	574.6561	784.7399	18.18	11/21 years
HIV-AIDS (NYC, 1982-2002)	2.7 (1.6, 4.4)	0.47 (0.33, 0.6)	6.1496	22.2194	81.81	11/21 years
Measles (London, 1948)	1.7 (1.3, 2.2)	0.51 (0.48, 0.55)	85.8138	135.7374	44.44	9/40 weeks
Pandemic influenza (San Fran, 1918)	0.29 (0.28, 0.34)	0.99 (0.95, 1)	8.4321	57.9749	60.00	20/63 days
Plague (Bombay, 1905-06)	0.12 (0.074, 0.17)	0.88 (0.79, 0.98)	6.5659	5.30811	100.00	9/41 weeks
Plague (Madagascar-wave2, 2017)	0.12 (0.072, 0.18)	0.81 (0.71, 0.94)	3.451	8.1155	100.00	11/50 weeks
Smallpox (Khulna, Bangladesh, 1972)	0.16 (0.11, 0.21)	0.85 (0.78, 0.93)	15.2023	17.3744	77.78	9/13 weeks

Table 2 Result of r and p parameters with 95% CI, RMSE, Anscombe residual, prediction coverage, and the length of ascending phase by MLE for each outbreak.

Outbreaks	r (95% CI)	p (95% CI)	RMSE	Anscombe	Prediction interval coverage (%)	length of ascending phase
Zika (Antioquia, 2016)	1.1 (0.82, 1.4)	0.55 (0.48, 0.61)	4.9396	25.6333	100.00	30 /104 days

Zika (Antioquia, 2016)	1.3 (0.75, 2.2)	0.49 (0.32, 0.65)	2.9777	15.5843	100.00	15/104 days
Zika (Antioquia, 2016)	1.2 (0.73, 2.1)	0.51 (0.34, 0.68)	3.6486	16.0586	100.00	16/104 days
FMD (UK,2001-120days)	0.5 (0.36, 0.67)	0.73 (0.64, 0.83)	4.005	37.2877	96.00	25/229 days
Ebola (Tonkolili, 2014)	0.15 (0.077, 0.32)	0.82 (0.58, 1)	6.5574	8.0175	83.33	6/69 weeks
Cholera (Aalborg,1853)	0.49 (0.34, 0.67)	0.81 (0.73, 0.89)	6.2008	36.4364	95.00	20/108 days
Ebola (Bo, 2014)	0.13 (0.089, 0.18)	0.81 (0.71, 0.92)	9.7877	27.4529	80.00	10/67 weeks
Ebola (Bombali, 2014)	0.075(0.063, 0.12)	0.97 (0.82, 1)	3.8406	16.0682	87.50	8/64 weeks
Ebola (Bomi, 2014)	1.1 (0.47, 2)	0.15 (0, 0.38)	6.8829	19.6677	75.00	8/66 weeks
Ebola (Congo, 1976)	1.1 (0.67, 1.8)	0.48, (0.32, 0.62)	3.4132	19.3494	95.00	20/52 days
Ebola (Grand Bassa, 2014)	0.35 (0.14, 0.72)	0.4 (0.1, 0.68)	3.7712	7.4662	100.00	9/64 weeks
Ebola (Gueckedou, 2014)	0.12 (0.042, 0.27)	0.69 (0.4, 0.97)	4	17.9151	90.91	11/90 weeks
Ebola (Kenema, 2014)	0.53 (0.34, 0.78)	0.49 (0.38, 0.6)	5.4314	17.3496	87.50	8/70 weeks
Ebola (Margibi, 2014)	0.13 90.11, 0.17)	0.86 (0.79, 0.93)	21.5708	57.1537	60.00	10/68 weeks
Ebola (Margibi, 2014)	0.15 (0.13, 0.19)	0.84 (0.75, 0.87)	16.8415	63.2808	54.55	11/68 weeks
Ebola (Montserrado, 2014)	0.15 (0.12, 0.2)	0.8 (0.72, 0.88)	9.7005	29.4463	70.00	10/71 weeks
Ebola (Port Loko, 2014)	0.55 (0.37, 0.8)	0.51 (0.41, 0.6)	4.8348	2.8323	100.00	8/64 weeks
Ebola (Uganda, 2000)	0.4 (0.25, 0.59)	0.62 (0.49, 0.76)	8.1138	1.5743	100.00	6/18 weeks
Ebola (Western Area Rural, 2014)	0.32 (0.24, 0.43)	0.62 (0.54, 0.69)	8.3666	12.5039	100.00	10/63 weeks
Ebola (Western Area Urban, 2014)	0.53 (0.36, 0.74)	0.52 (0.44, 0.6)	6.9785	12.0879	90.00	10/62 weeks
FMD (Uruguay, 2001)	8.2 (6.7, 9.9)	0.54 (0.51, 0.57)	62.4915	161.757	36.36	11/27 days
HIV-AIDS (Japan, 1985-2012)	0.081 (0.076, 0.086)	0.53 (0.52, 0.53)	551.8599	725.4638	18.18	11/21 years
HIV-AIDS (NYC, 1982-2002)	2.3 (1.4, 3.4)	0.51 (0.4, 0.63)	4.5826	21.4924	81.81	11/21 years
Measles (London, 1948)	2.9 (2.3, 3.6)	0.44 (0.41, 0.47)	67.5681	118.335	33.33	9/40 weeks
Pandemic influenza (San Fran, 1918)	0.35 (0.3, 0.41)	0.94 (0.91, 0.98)	14.0535	52.5996	70.00	20/63 days
Plague (Bombay, 1905-06)	0.12 (0.085, 0.17)	0.88 (0.78, 0.98)	4.3205	4.9868	100.00	9/41 weeks
Plague (Madagascar-wave2, 2017)	0.11 (0.073, 0.15)	0.84 (0.75, 0.93)	6.6332	7.5833	100.00	11/50 weeks
Smallpox (Khulna, Bangladesh, 1972)	0.14 (0.11, 0.17)	0.87 (0.82, 0.93)	15.8184	16.3525	88.89	9/13 weeks

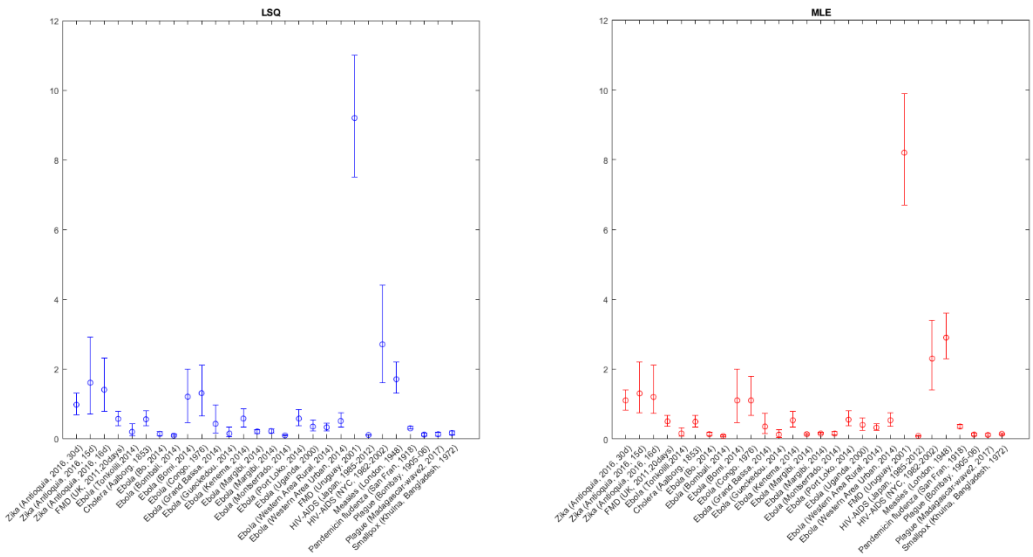
Table 3 Difference between LSQ and MLE results. (LSQ-MLE)

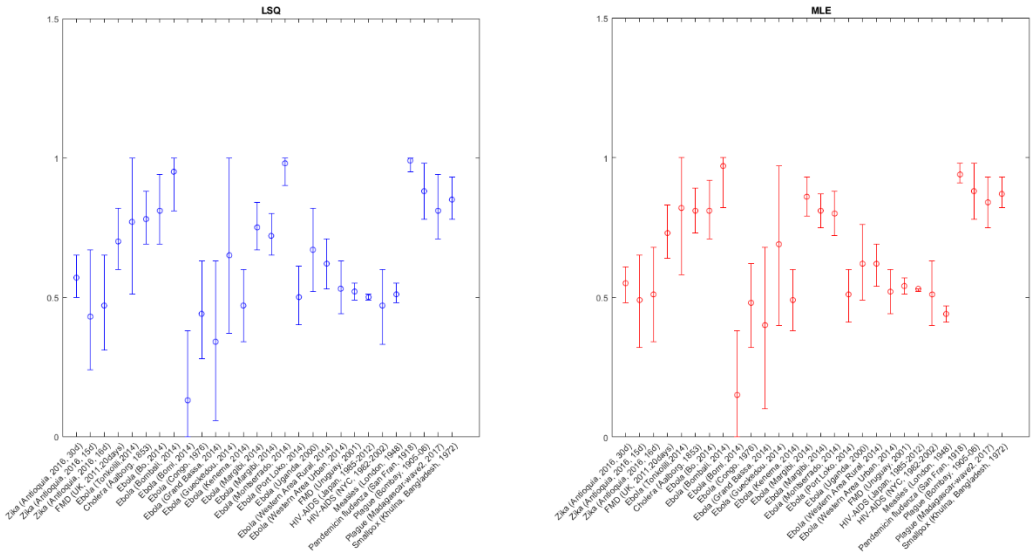
Difference	r	p	RMSE	Abs. RMSE	Anscombe	Abs. Anscombe	p_coverage (%)	Abs. value of p_coverage (%)	length of ascending phase
Zika (Antioquia, 2016)	-0.13	0.02	-0.2849	0.2849	0.5368	0.5368	-3.33	3.33	30 /104 days
Zika (Antioquia, 2016)	0.3	-0.06	0.7371	0.7371	0.7171	0.7171	0.00	0.00	15/104 days
Zika (Antioquia, 2016)	0.2	-0.04	-0.3414	0.3414	0.3469	0.3469	0.00	0.00	16/104 days
FMD (UK,2001-120days)	0.06	-0.03	-0.7492	0.7492	0.79	0.79	0.00	0.00	25/229 days
Ebola (Tonkolili, 2014)	0.04	-0.05	-1.814	1.814	0.2389	0.2389	0.00	0.00	6/69 weeks
Cholera (Aalborg,1853)	0.06	-0.03	1.6573	1.6573	0.3258	0.3258	-5.00	5.00	20/108 days
Ebola (Bo, 2014)	0	0	-0.1285	0.1285	-0.0084	0.0084	0.00	0.00	10/67 weeks
Ebola (Bombali, 2014)	0.006	-0.02	0.2368	0.2368	0.9929	0.9929	0.00	0.00	8/64 weeks
Ebola (Bomi, 2014)	0.1	-0.02	-0.9457	0.9457	-0.002	0.002	0.00	0.00	8/66 weeks

Ebola (Congo, 1976)	0.2	-0.04	0.2539	0.2539	0.2434	0.2434	5.00	5.00	20/52 days
Ebola (Grand Bassa, 2014)	0.07	-0.06	-0.2593	0.2593	0.2312	0.2312	0.00	0.00	9/64 weeks
Ebola (Gueckedou, 2014)	0.02	-0.04	-0.1506	0.1506	0.098	0.098	0.00	0.00	11/90 weeks
Ebola (Kenema, 2014)	0.04	-0.02	3.7064	3.7064	0.1982	0.1982	0.00	0.00	8/70 weeks
Ebola (Margibi, 2014)	0.07	-0.11	-5.8069	5.8069	10.6799	10.6799	-10.00	10.00	10/68 weeks
Ebola (Margibi, 2014)	0.06	-0.09	1.1256	1.1256	9.9955	9.9955	0.00	0.00	11/68 weeks
Ebola (Montserrado, 2014)	-0.062	0.18	-0.3357	0.3357	17.8647	17.8647	-20.00	20.00	10/71 weeks
Ebola (Port Loko, 2014)	0.02	-0.01	4.0534	4.0534	0.0312	0.0312	0.00	0.00	8/64 weeks
Ebola (Uganda, 2000)	-0.06	0.05	-2.966	2.966	0.4052	0.4052	0.00	0.00	6/18 weeks
Ebola (Western Area Rural, 2014)	0	0	-1.3026	1.3026	0.0025	0.0025	-10.00	10.00	10/63 weeks
Ebola (Western Area Urban, 2014)	-0.03	0.01	0.1278	0.1278	0.0549	0.0549	0.00	0.00	10/62 weeks
FMD (Uruguay, 2001)	1	-0.02	15.6901	15.6901	1.6608	1.6608	0.00	0.00	11/27 days
HIV-AIDS (Japan, 1985-2012)	0.019	-0.03	22.7962	22.7962	59.2761	59.2761	0.00	0.00	11/21 years
HIV-AIDS (NYC, 1982-2002)	0.4	-0.04	1.567	1.567	0.727	0.727	0.00	0.00	11/21 years
Measles (London, 1948)	-1.2	0.07	18.2457	18.2457	17.4024	17.4024	11.11	11.11	9/40 weeks
Pandemic influenza (San Fran, 1918)	-0.06	0.05	-5.6214	5.6214	5.3753	5.3753	-10.00	10.00	20/63 days
Plague (Bombay, 1905-06)	0	0	2.2454	2.2454	0.32131	0.32131	0.00	0.00	9/41 weeks
Plague (Madagascar-wave2, 2017)	0.01	-0.03	-3.1822	3.1822	0.5322	0.5322	0.00	0.00	11/50 weeks
Smallpox (Khulna, Bangladesh, 1972)	0.02	-0.02	-0.6161	0.6161	1.0219	1.0219	-11.11	11.11	9/13 weeks

List of Figures

(Use this page only if you have included figures in your thesis/capstone.)





Systematic comparison of parameter estimation approaches using the generalized-growth model for prediction of epidemic outbreaks

Yiseul Lee[1*], Kimberlyn Roosa[1], Amna Tariq[1], and Gerardo Chowell[1, 2]
[1] Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA
[2] Division of International Epidemiology and Population Studies, Fogarty International Center, National Institute of Health, Bethesda, MD, USA

*__Correspondence__: Email: ylee97@student.gsu.edu

Abstract

Background- Many different mathematical models are used to assess and predict the outbreaks. The model is selected by the characteristics of the outbreaks. Here, we utilize the generalized growth model (GGM), one of the simplest mathematical models, with the real outbreaks to compare two parameter estimation methods.
Materials and Methods- 25 outbreaks are used to analyze. We use GGM with the ascending phase of each outbreak and estimated the r and p parameters with both the least square (LSQ) and maximum likelihood estimation (MLE) methods. For both parameter estimation methods, we conduct the parametric bootstrap method to construct the confidence interval of parameters. We compare the two estimation methods by the RMSE, Anscombe residual, and prediction coverage.
Results- The result show that the most outbreaks have similar r and p parameters, RMSE, Anscombe, and prediction coverage for LSQ and MLE. Although Anscombe values for LSQ are higher than the values for MLE, the different between results of the two methods are minimal for the most outbreaks.
Conclusion- The study is shown that LSQ and MLE do not result in different values of the parameter estimation, RMSE, Anscombe, and prediction coverage with GGM.

Keywords: parameter estimation; generalized growth model; least square estimation, maximum likelihood estimation; epidemiological models

1. Introduction

Mathematical models are utilized in infectious disease epidemiology to assess, investigate and forecast epidemic outbreaks and understand the underlying mechanisms of disease transmission. These mathematical model frameworks also help us understand the epidemic control strategies. Various mathematical models are used to investigate epidemic outbreaks, including GGM and generalized logistic model (GLM), with each model defined by the characteristics of the outbreak. One of the defining outbreak characteristics is the epidemic growth pattern. Epidemic growth patterns vary substantially depending on distinctive features of the outbreaks. Even within the same type of disease, outbreak patterns can differ, as can be observed with influenza virus disease (1). Many infectious disease studies, including studies on the influenza virus epidemics, begin with the assumption of an exponential epidemic growth pattern. As has been observed historically, influenza outbreaks tend to follow an exponential growth pattern. The virus spreads quickly via coughing and droplet spread through air and exhibits a low fatality rate

(0.9%); however, the fatality rate could vary depending on the genotype of the influenza virus circulating in the air (2). The exponential growth assumption is not always valid, as infectious diseases have other distinct features such as disease transmission methods and incubation period, which influence the epidemic disease growth pattern. For example, Ebola is a viral disease like influenza, but it spreads via close contact with an Ebola-infected patient and has a high fatality rate. Another example of a retroviral disease with high fatality rate is the human immunodeficiency virus disease (HIV), which also spreads via close contact with an HIV infected patient (4). Since both the aforementioned diseases transmit via close contact, their rate of spread is slower compared to influenza. For such diseases with lower rate of spread, sub-exponential growth patterns would be expected (3). Further, population structure, environmental factors, vaccination policies, etc. also impact the spread of disease within a population.

Several methods have been proposed to estimate the parameters in a model; these include least square estimation (LSQ) and maximum likelihood estimation (MLE) methods. The choice of method employed to estimate model parameters depends on the characteristics of the model and the available data. In this paper, both the LSQ and MLE methods are utilized to estimate model parameters assuming Poisson error structure like works in the previous paper (5). Previous work by Roosa et al., (2019) shows that LSQ with parametric bootstrapping and MLE assuming Poisson distribution yielded very similar results utilizing data simulated from the model (Roosa et al., 2019) (5). In their paper, they compare the r and p parameters of GGM and GLM for LSQ and MLE methods with the simulated data (5). Based on the results from their paper, for the purpose of our analysis we further evaluate the GGM for the LSQ and MLE methods using data from real disease outbreaks. We use the parametric bootstrapping method to predict the best guess of the initial parameters and assess the variables using the best guess of the initial parameters.

2. Data and Methods

2.1. *Data sources*

Data from 25 different outbreaks are used for the purpose of this analysis. Outbreak data consists of Zika (Antioqia, 2016, 30, 15 & 16 days), FMD (UK, 2001, 25 days), Ebola (Tonkolili, 2014, 6 weeks), Cholera (Aalborg, 1853, 20 days), Ebola (Bo, 2014, 10 weeks), Ebola (Bombali, 2014, 8 weeks), Ebola (Bomi, 2014, 8 weeks), Ebola (Congo, 1976, 20 days), Ebola (Grand Bassa, 2014, 9 weeks), Ebola (Gueckedou, 2014, 11 weeks), Ebola (Kenema, 2014, 8 weeks), Ebola (Margibi, 2014, 10 & 11 weeks), Ebola (Montserrado, 2014, 10 weeks), Ebola (Port Loko, 2014, 8 weeks), Ebola (Uganda, 2000, 6 weeks), Ebola (Western Area Rural, 2014, 10 weeks), Ebola (Western Area Urban, 2014, 10 weeks), FMD (Urugay, 2001, 11 days), HIV-AIDS (Japan, 1985-2012, 11 years), HIV-AIDS (NYC, 1982-2002, 11 years), Measles (London, 1948, 9 weeks), Pandemic Influenza (San Francisco, 1918, 20 days), Plague (Bombay, 1905-06, 9 weeks), Plague (Madagascar-wave2, 2017, 11 weeks), and Smallpox (Khulna, Bangladesh, 1972, 9 weeks) (Table 1-1). The temporal scale of the data varies depending on the outbreaks, with most outbreaks reported daily or weekly. However, the HIV-AIDS outbreaks have year as the unit of analysis. Therefore, in order to standardize the unit of time for our data analysis, we assume day as the unit of time. Hence, results depict day as our final unit of time.

The data is named in order disease, the place, year or period it happened and the length of ascending phase, or the number of data points we use. Different lengths of ascending phase are used for two of the outbreaks: 30, 15 and 16 data points for Zika (Antioqia, 2016) and 10 and 11 data points for Ebola (Margibi, 2014).

2.2.1. *Generalized Growth Model (GGM)*

As we explain in the introduction, GGM allows for less than exponential growth patterns. The generalized growth model is composed of two parameters and one variable. The two parameters include the "deceleration growth parameter" p and the growth rate parameter, r. C (t) represents the cumulative number of cases at time t and C′(t) represents the incidence curve. When the "deceleration of growth" parameter (p) lies within the range of 0 and 1 it depicts the sub-exponential growth patterns, p=0 shows constant/linear growth, whereas p=1 shows an exponential pattern. Growth rate, r parameter, is greater than 0 (6).

The GGM equation is the following:

$$\frac{dC(t)}{dt} = C'(t) = rC(t)^p$$

Various outbreaks including Zika (7, 10), Foot and Mouth disease (8), Ebola (9), and HIV- AIDS (11) have been modelled using the GGM.

2.3. *Parameter estimation*

For assessing the parameters, we conduct parametric bootstrap analyses using LSQ and MLE methods. A previous study shows that one can evaluate parameter with a simple computational bootstrap-based method, by replicating several data sets through repeated sampling from the model. (17). When estimating parameters, the initial parameter values or "guesses" can impact the results due to local maxima or minima. Therefore, we utilize the bootstrap method several times with different initial parameter guesses to estimate the best initial parameters, or those with the lowest MSE, for the 'best-fit' model. With the initial parameters from the bootstrap method, we then use these values and employ the bootstrapping method to simulate 500 curves (M=500) from the best-fit model, and further, re-estimate the parameters for each of these new datasets. We then utilize the parameter estimate distributions to calculate 95% confidential intervals, root mean square error, prediction intervals, and Anscombe residuals.

2.3.1. *Least squares estimation (LSQ)*

The goal of the least square estimation method is to estimate the best fit line for the data. This method yields the best fit solution for the given time by exploring the parameters while minimizing the sum of the squared deviations between the data and the model solution. The equation is follows:

$$\hat{\theta} = argmin \sum_{t=1}^{n} (f(t;\theta) - y_t)^2$$

$\hat{\theta}$ is representative of the parameters r and p employed in the GGM, y_t represents the outbreaks' data, and f (t; $\hat{\theta}$) =C'(t|$\hat{\theta}$) is the incidence cases in the GGM model. We used *fmincon* function in Matlab 2017 to get the nonlinear least squares estimation for our model parameters and restrict the bounds for each parameter. With this method, the overall data of the parameter has equal weight (13). We employ the bootstrap method to quantify the uncertainty of the parameter estimates. The method of LSQ with bootstrap approach is as follows:

1. Estimate the best initial parameters.
2. With the best initial parameters, fit the model to the time series data and derive parameter estimates with LSQ.
3. Generate M=500 outbreaks data sets. For each data set, assume Poisson error distribution with mean=C'(t), which is the incidence curve for the time t.
4. With each of the new outbreak data sets, estimate the r and p parameters with LSQ, prediction interval, and Anscombe residual.
5. Based on the 500 estimated values of parameters, construct the 95% confidential interval and calculate the RMSE.

2.3.2. *Maximum likelihood estimation (MLE)*

The maximum likelihood estimation method estimates the parameters that maximize the likelihood function of the data. MLE has the same goal as LSQ in terms of estimating the parameters, but the primary difference between LSQ and MLE lies in the utilization of weighted points. LSQ puts equal weight for all data points, while MLE puts different weight for the data point depending on the magnitude (13, 14).

The equation of the MLE estimate is:

$$\hat{\theta} = argmax \sum_{t=1}^{n} [y_t log\big(f(t;\theta)\big) - f(t;\theta)]$$

The method of MLE with bootstrap approach is as follows:

1. Estimate the best initial parameters.
2. With the best initial parameters, fit the model to the time series data and derive parameter estimates with MLE.
3. Generate M=500 outbreaks data sets. For each data set, assume Poisson error distribution with mean=C'(t), which is the incidence curve for the time t.
4. With each of the new outbreak data sets, estimate the r and p parameters with MLE, prediction interval, and Anscombe residual.
5. Based on the 500 estimated values of parameters, construct the 95% confidential interval and calculate the RMSE.

The variable expression is the same as with LSQ. We also employed the *fmincon* function in Matlab 2017 like LSQ. For the purpose of this research paper, we compare the results of the LSQ and MLE with real outbreaks.

2.4. *Root Mean Square Error (RMSE) and Anscombe residual*

The residual represents the difference between the true parameter value and the distribution of parameter estimates. The residual shows the deviation of the model fit from the data and assesses the performance of the method for the model with the data (15). One widely used metric is Mean square Error (MSE), along with the widely used RMSE. The equations for these are as follows:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}[y_t - f(t;\theta)]^2$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}[y_t - f(t;\theta)]^2}$$

To account for individual weights of the data points, we use the Anscombe residual depicted in the book by McCullagh and Melder (1983) (16) which has equation as follows:

$$ANSCOMBE\ RESIDUAL = \frac{\frac{3}{2}\left[yt^{2/3} - f(t;\theta)^{2/3}\right]}{f(t;\theta)^{1/6}}$$

$$ANSCOMBE = \sum_{t=1}^{T}\left(\frac{\frac{3}{2}\left[y_t{}^{2/3} - f(t;\theta)^{2/3}\right]}{f(t;\theta)^{1/6}}\right)^2$$

2.5. *Performance*

From each LSQ and MLE method, estimates of parameters r and p along with 95% CI, root mean square error (RMSE), Anscombe and prediction coverage were calculated for the given outbreaks. For assessing the difference between LSQ and MLE results, we subtracted each variable, r and p parameters, RMSE, Anscombe, 95% prediction interval coverage, from LSQ to MLE (LSQ-MLE) and calculated the absolute values.

3. Results

Table 1 Result of r and p parameters with 95% CI, RMSE, Anscombe residual, prediction coverage, and the length of ascending phase by LSQ for each outbreak.

LSQ	r (95% CI)	p (95% CI)	RMSE	Anscombe	Prediction interval coverage (%)	length of ascending phase
Zika (Antioquia, 2016)	0.97 (0.69, 1.3)	0.57 (0.5, 0.65)	4.6547	26.1701	96.67	30 /104 days
Zika (Antioquia, 2016)	1.6 (0.71, 2.9)	0.43 (0.24, 0.67)	3.7148	16.3014	100.00	15/104 days
Zika (Antioquia, 2016)	1.4 (0.78, 2.3)	0.47 (0.31, 0.65)	3.3072	16.4055	100.00	16/104 days

FMD (UK,2001-120days)	0.56 (0.37, 0.78)	0.7 (0.6, 0.82)	3.2558	38.0777	96.00	25/229 days
Ebola (Tonkolili, 2014)	0.19 (0.079, 0.42)	0.77 (0.51, 1)	4.7434	8.2564	83.33	6/69 weeks
Cholera (Aalborg,1853)	0.55 (0.36, 0.8)	0.78 (0.69, 0.88)	7.8581	36.7622	90.00	20/108 days
Ebola (Bo, 2014)	0.13 (0.083, 0.2)	0.81 (0.69, 0.94)	9.6592	27.4445	80.00	10/67 weeks
Ebola (Bombali, 2014)	0.081 (0.064, 0.13)	0.95 (0.81, 1)	4.0774	17.0611	87.50	8/64 weeks
Ebola (Bomi, 2014)	1.2 (0.46, 2)	0.13 (0, 0.38)	5.9372	19.6657	75.00	8/66 weeks
Ebola (Congo, 1976)	1.3 (0.65, 2.1)	0.44 (0.28, 0.63)	3.6671	19.5928	100.00	20/52 days
Ebola (Grand Bassa, 2014)	0.42 (0.15, 0.97)	0.34 (0.058, 0.63)	3.5119	7.6974	100.00	9/64 weeks
Ebola (Gueckedou, 2014)	0.14 (0.04, 0.32)	0.65 (0.37, 1)	3.8494	18.0131	90.91	11/90 weeks
Ebola (Kenema, 2014)	0.57 (0.33, 0.85)	0.47 (0.34, 0.6)	9.1378	17.5478	87.50	8/70 weeks
Ebola (Margibi, 2014)	0.2 (0.14, 0.26)	0.75 (0.67, 0.84)	15.7639	67.8336	50.00	10/68 weeks
Ebola (Margibi, 2014)	0.21 (0.16, 0.28)	0.72 (0.65, 0.8)	17.9671	73.2763	54.55	11/68 weeks
Ebola (Montserrado, 2014)	0.088 (0.081, 0.11)	0.98 (0.9, 1)	9.3648	47.311	50.00	10/71 weeks
Ebola (Port Loko, 2014)	0.57 (0.36, 0.83)	0.5 (0.4, 0.61)	8.8882	2.8635	100.00	8/64 weeks
Ebola (Uganda, 2000)	0.34 (0.21, 0.53)	0.67 (0.52, 0.82)	5.1478	1.9795	100.00	6/18 weeks
Ebola (Western Area Rural, 2014)	0.32 (0.22, 0.44)	0.62 (0.53, 0.71)	7.064	12.5064	90.00	10/63 weeks
Ebola (Western Area Urban, 2014)	0.5 (0.32, 0.73)	0.53 (0.44, 0.63)	7.1063	12.1428	90.00	10/62 weeks
FMD (Uruguay, 2001)	9.2 (7.5, 11)	0.52 (0.49, 0.55)	78.1816	163.4178	36.36	11/27 days
HIV-AIDS (Japan, 1985-2012)	0.1 (0.097, 0.11)	0.5 (0.49, 0.51)	574.6561	784.7399	18.18	11/21 years
HIV-AIDS (NYC, 1982-2002)	2.7 (1.6, 4.4)	0.47 (0.33, 0.6)	6.1496	22.2194	81.81	11/21 years
Measles (London, 1948)	1.7 (1.3, 2.2)	0.51 (0.48, 0.55)	85.8138	135.7374	44.44	9/40 weeks
Pandemic influenza (San Fran, 1918)	0.29 (0.28, 0.34)	0.99 (0.95, 1)	8.4321	57.9749	60.00	20/63 days
Plague (Bombay, 1905-06)	0.12 (0.074, 0.17)	0.88 (0.79, 0.98)	6.5659	5.30811	100.00	9/41 weeks
Plague (Madagascar-wave2, 2017)	0.12 (0.072, 0.18)	0.81 (0.71, 0.94)	3.451	8.1155	100.00	11/50 weeks
Smallpox (Khulna, Bangladesh, 1972)	0.16 (0.11, 0.21)	0.85 (0.78, 0.93)	15.2023	17.3744	77.78	9/13 weeks

Table 2 Result of r and p parameters with 95% CI, RMSE, Anscombe residual, prediction coverage, and the length of ascending phase by MLE for each outbreak.

Outbreaks	r (95% CI)	p (95% CI)	RMSE	Anscombe	Prediction interval coverage (%)	length of ascending phase
Zika (Antioquia, 2016)	1.1 (0.82, 1.4)	0.55 (0.48, 0.61)	4.9396	25.6333	100.00	30 /104 days
Zika (Antioquia, 2016)	1.3 (0.75, 2.2)	0.49 (0.32, 0.65)	2.9777	15.5843	100.00	15/104 days
Zika (Antioquia, 2016)	1.2 (0.73, 2.1)	0.51 (0.34, 0.68)	3.6486	16.0586	100.00	16/104 days
FMD (UK,2001-120days)	0.5 (0.36, 0.67)	0.73 (0.64, 0.83)	4.005	37.2877	96.00	25/229 days
Ebola (Tonkolili, 2014)	0.15 (0.077, 0.32)	0.82 (0.58, 1)	6.5574	8.0175	83.33	6/69 weeks
Cholera (Aalborg,1853)	0.49 (0.34, 0.67)	0.81 (0.73, 0.89)	6.2008	36.4364	95.00	20/108 days
Ebola (Bo, 2014)	0.13 (0.089, 0.18)	0.81 (0.71, 0.92)	9.7877	27.4529	80.00	10/67 weeks
Ebola (Bombali, 2014)	0.075(0.063, 0.12)	0.97 (0.82, 1)	3.8406	16.0682	87.50	8/64 weeks
Ebola (Bomi, 2014)	1.1 (0.47, 2)	0.15 (0, 0.38)	6.8829	19.6677	75.00	8/66 weeks
Ebola (Congo, 1976)	1.1 (0.67, 1.8)	0.48, (0.32, 0.62)	3.4132	19.3494	95.00	20/52 days
Ebola (Grand Bassa, 2014)	0.35 (0.14, 0.72)	0.4 (0.1, 0.68)	3.7712	7.4662	100.00	9/64 weeks

Ebola (Gueckedou, 2014)	0.12 (0.042, 0.27)	0.69 (0.4, 0.97)	4	17.9151	90.91	11/90 weeks
Ebola (Kenema, 2014)	0.53 (0.34, 0.78)	0.49 (0.38, 0.6)	5.4314	17.3496	87.50	8/70 weeks
Ebola (Margibi, 2014)	0.13 90.11, 0.17)	0.86 (0.79, 0.93)	21.5708	57.1537	60.00	10/68 weeks
Ebola (Margibi, 2014)	0.15 (0.13, 0.19)	0.84 (0.75, 0.87)	16.8415	63.2808	54.55	11/68 weeks
Ebola (Montserrado, 2014)	0.15 (0.12, 0.2)	0.8 (0.72, 0.88)	9.7005	29.4463	70.00	10/71 weeks
Ebola (Port Loko, 2014)	0.55 (0.37, 0.8)	0.51 (0.41, 0.6)	4.8348	2.8323	100.00	8/64 weeks
Ebola (Uganda, 2000)	0.4 (0.25, 0.59)	0.62 (0.49, 0.76)	8.1138	1.5743	100.00	6/18 weeks
Ebola (Western Area Rural, 2014)	0.32 (0.24, 0.43)	0.62 (0.54, 0.69)	8.3666	12.5039	100.00	10/63 weeks
Ebola (Western Area Urban, 2014)	0.53 (0.36, 0.74)	0.52 (0.44, 0.6)	6.9785	12.0879	90.00	10/62 weeks
FMD (Uruguay, 2001)	8.2 (6.7, 9.9)	0.54 (0.51, 0.57)	62.4915	161.757	36.36	11/27 days
HIV-AIDS (Japan, 1985-2012)	0.081 (0.076, 0.086)	0.53 (0.52, 0.53)	551.8599	725.4638	18.18	11/21 years
HIV-AIDS (NYC, 1982-2002)	2.3 (1.4, 3.4)	0.51 (0.4, 0.63)	4.5826	21.4924	81.81	11/21 years
Measles (London, 1948)	2.9 (2.3, 3.6)	0.44 (0.41, 0.47)	67.5681	118.335	33.33	9/40 weeks
Pandemic influenza (San Fran, 1918)	0.35 (0.3, 0.41)	0.94 (0.91, 0.98)	14.0535	52.5996	70.00	20/63 days
Plague (Bombay, 1905-06)	0.12 (0.085, 0.17)	0.88 (0.78, 0.98)	4.3205	4.9868	100.00	9/41 weeks
Plague (Madagascar-wave2, 2017)	0.11 (0.073, 0.15)	0.84 (0.75, 0.93)	6.6332	7.5833	100.00	11/50 weeks
Smallpox (Khulna, Bangladesh, 1972)	0.14 (0.11, 0.17)	0.87 (0.82, 0.93)	15.8184	16.3525	88.89	9/13 weeks

Table 3 Difference between LSQ and MLE results. (LSQ-MLE)

Difference	r	p	RMSE	Abs. RMSE	Anscombe	Abs. Anscombe	Prediction interval coverage (%)	Abs. value of p_coverage (%)	length of ascending phase
Zika (Antioquia, 2016)	-0.13	0.02	-0.2849	0.2849	0.5368	0.5368	-3.33	3.33	30 /104 days
Zika (Antioquia, 2016)	0.3	-0.06	0.7371	0.7371	0.7171	0.7171	0.00	0.00	15/104 days
Zika (Antioquia, 2016)	0.2	-0.04	-0.3414	0.3414	0.3469	0.3469	0.00	0.00	16/104 days
FMD (UK,2001-120days)	0.06	-0.03	-0.7492	0.7492	0.79	0.79	0.00	0.00	25/229 days
Ebola (Tonkolili, 2014)	0.04	-0.05	-1.814	1.814	0.2389	0.2389	0.00	0.00	6/69 weeks
Cholera (Aalborg,1853)	0.06	-0.03	1.6573	1.6573	0.3258	0.3258	-5.00	5.00	20/108 days
Ebola (Bo, 2014)	0	0	-0.1285	0.1285	-0.0084	0.0084	0.00	0.00	10/67 weeks
Ebola (Bombali, 2014)	0.006	-0.02	0.2368	0.2368	0.9929	0.9929	0.00	0.00	8/64 weeks
Ebola (Bomi, 2014)	0.1	-0.02	-0.9457	0.9457	-0.002	0.002	0.00	0.00	8/66 weeks
Ebola (Congo, 1976)	0.2	-0.04	0.2539	0.2539	0.2434	0.2434	5.00	5.00	20/52 days
Ebola (Grand Bassa, 2014)	0.07	-0.06	-0.2593	0.2593	0.2312	0.2312	0.00	0.00	9/64 weeks
Ebola (Gueckedou, 2014)	0.02	-0.04	-0.1506	0.1506	0.098	0.098	0.00	0.00	11/90 weeks
Ebola (Kenema, 2014)	0.04	-0.02	3.7064	3.7064	0.1982	0.1982	0.00	0.00	8/70 weeks
Ebola (Margibi, 2014)	0.07	-0.11	-5.8069	5.8069	10.6799	10.6799	-10.00	10.00	10/68 weeks
Ebola (Margibi, 2014)	0.06	-0.09	1.1256	1.1256	9.9955	9.9955	0.00	0.00	11/68 weeks

Ebola (Montserrado, 2014)	-0.062	0.18	-0.3357	0.3357	17.8647	17.8647	-20.00	20.00	10/71 weeks
Ebola (Port Loko, 2014)	0.02	-0.01	4.0534	4.0534	0.0312	0.0312	0.00	0.00	8/64 weeks
Ebola (Uganda, 2000)	-0.06	0.05	-2.966	2.966	0.4052	0.4052	0.00	0.00	6/18 weeks
Ebola (Western Area Rural, 2014)	0	0	-1.3026	1.3026	0.0025	0.0025	-10.00	10.00	10/63 weeks
Ebola (Western Area Urban, 2014)	-0.03	0.01	0.1278	0.1278	0.0549	0.0549	0.00	0.00	10/62 weeks
FMD (Uruguay, 2001)	1	-0.02	15.6901	15.6901	1.6608	1.6608	0.00	0.00	11/27 days
HIV-AIDS (Japan, 1985-2012)	0.019	-0.03	22.7962	22.7962	59.2761	59.2761	0.00	0.00	11/21 years
HIV-AIDS (NYC, 1982-2002)	0.4	-0.04	1.567	1.567	0.727	0.727	0.00	0.00	11/21 years
Measles (London, 1948)	-1.2	0.07	18.2457	18.2457	17.4024	17.4024	11.11	11.11	9/40 weeks
Pandemic influenza (San Fran, 1918)	-0.06	0.05	-5.6214	5.6214	5.3753	5.3753	-10.00	10.00	20/63 days
Plague (Bombay, 1905-06)	0	0	2.2454	2.2454	0.32131	0.32131	0.00	0.00	9/41 weeks
Plague (Madagascar-wave2, 2017)	0.01	-0.03	-3.1822	3.1822	0.5322	0.5322	0.00	0.00	11/50 weeks
Smallpox (Khulna, Bangladesh, 1972)	0.02	-0.02	-0.6161	0.6161	1.0219	1.0219	-11.11	11.11	9/13 weeks

For each outbreak, the results with graphs including r and p parameters with 95% confidence interval, GGM with prediction interval, and Anscombe residual are at the appendix part (Figure S1-1 & S1-2). For r and p parameters, the red line represents to the mean with bootstrapping and red dot is 95% confidence interval. The left lower graph represents fitting the model with data. The blue bubble represents the data, the red line is best fit model, cyan lines represent the 500 bootstrapping with GGM, and the red dot is for 95% prediction interval. The right lower graph is Anscombe residual (Appendix. Figure S1-1 & S1-2).

The majority of the of 25 outbreaks that are analyzed for the purpose of this analysis had similar results for the growth parameter (r) (CI 95%), the deceleration parameter (p) (CI 95%), RMSE, Anscombe residual, and the prediction interval with LSQ and MLE methods. Growth parameter r and Anscombe residual have higher values for the LSQ method for most outbreaks. On the other hand, parameter p, shows higher value with the MLE method for most outbreaks. Values of RMSE and prediction coverage for the outbreaks have almost similar values with LSQ and MLE methods.

R parameter for most outbreaks with LSQ method shows equal with or higher than than MLE method except 6 outbreaks, Zika (Antioquia, 2016, 30 days), Ebola (Montserrado, 2014), Ebola (Uganda, 2000), Ebola (Western Area Urban, 2014), Measles (London, 1948), and Pandemic influenza (San Fran, 1918). The range of difference r parameter with the absolute r parameter value is from 0 to 1.2. The outbreaks which has the most different r parameter is Measles (London, 1948) with 1.2, 1.7 with LSQ and 2.9 with MLE (Table 3 & Figure 1-1).

The differences between LSQ and MLE for the p parameters of the outbreaks show that most outbreaks have higher p parameter when estimated using MLE. 9 outbreaks show that equal or high value when estimated using LSQ. The range of difference p parameter with the absolute p parameter value is from 0 to 0.18. The outbreaks which has the most different p parameter is Ebola (Montserrado, 2014) with 0.18, 0.98 with LSQ and 0.8 with LSQ (Table 3 & Figure 1-2).

RMSE and Anscombe's residual show similar results. The range (minimum to maximum) of Anscombe's residual with the LSQ method is observed to be between 3.2558 for Foot and Mouth Disease (UK, 2001) and 574.6561 for HIV-AIDS epidemic (Japan, 1985-2012) .Whereas, the range of RMSE with the MLE method is between 2.9777 (as observed for the Zika epidemic (Antioquia, 2016) with 15 data points) and 551.8599 (as observed with the HIV-AIDS epidemic (Japan, 1985-2012)). The range of the difference between two methods of the RMSE is between 0.1506 and 22.7962. The lowest difference value is Ebola (Gueckedou, 2014), and the highest difference value is HIV-AIDS (Japan, 1985-2012). 13 out of 28 outbreaks have higher RMSE for LSQ compared to MLE.

For the Anscombe residual, we observe a broad range of values for the outbreaks, ranging between 1.9795 and 784.7399 with LSQ method and between 1.5743 and 725.4638 with MLE. However, the difference of the Anscombe residual for the two methods ranges between 0.002 and 59.2761. The lowest difference value for the Anscombe residual is observed for the Ebola outbreak (Bomi, 2014) and the highest difference is observed for the HIV-AIDS outbreak (Japan, 1985-2012). For the majority of the outbreaks analyzed, the Anscombe residual estimated by the MLE method is lower than the Anscombe residual for the LSQ method, with the exception of two outbreaks where the Anscombe residual estimated by the LSQ method is lower than that estimated by the MLE method (Ebola (Bo, 2014) and Ebola (Bomi, 2014)).

Prediction interval coverage for most outbreaks with LSQ method show over 80% coverage, excluding 75% for Ebola (Bomi, 2014), 50% and 54.55% for Ebola (Margibi, 2014) using 10 and 11 data points respectively, 50% for Ebola (Montserrado, 2014), 36.36% for FMD (Uruguay, 2001), 44.44% for measles (London, 1948), 60% for Pandemic influenza (San Fransisco, 1918), and 77.78% for smallpox (Khulna, Bangladesh, 1972). Most values of the prediction inverval coverage with MLE method also show a prediction coverage of over 80%, excluding 75% coverage for Ebola (Bomi, 2014), 60% and 54.55% for Ebola (Margibi, 2014) for 10 and 11 data points respectively, 70% for Ebola (Montserrado, 2014), 36.36% for FMD (Uruguay, 2001), 18.18% for HIV-AIDS (Japan, 1985-2012), and 33.33% for measles (London, 1948). Between the two methods, we observe a difference of less than 5 % for the prediction coverage for most outbreaks. However, approximately 6 outbreaks have more than 5% difference in the prediction coverage between the two methods, and these include a 10% difference in the prediction coverage of two methods for the Ebola outbreak (Margibi, 2014) using 10 data points, 20% difference in the Ebola (Montserrado, 2014), 10% difference in the Ebola (Western Area, Rural, 2014), 11.11% difference in the measles (London, 1948), a 10% difference in the pandemic influenza (San Francisco, 1918), and an 11.11% difference in the smallpox (Khulna Bangladesh, 1972).

For the 95% prediction coverage, most of the outbreaks show the same coverage when estimated using LSQ and MLE methods. However, 9 outbreaks show a difference in the estimated values of 95% prediction coverage between the two methods. Among these 9 outbreaks, 7 outbreaks have a higher 95% prediction coverage when estimated using the MLE method compared to the coverage for the LSQ method. The most deviant result is observed for the Ebola (Montserrado, 2014) outbreak with a difference of 20% in the prediction coverage with 50% when estimated using LSQ method and 70% when estimated using MLE method.

Among 25 outbreaks, 14 outbreaks are Ebola. Among the Ebola outbreaks, the growth parameter, r, ranges from 0.08 for the LSQ (Ebola (Bombali, 2014)) to 1.3 (Ebola (Congo, 1976)), and for the MLE method the growth parameter ranges from 0.08 (Ebola (Bombali, 2014)) to 1.1 (Ebola (Bomi, 2014) and Ebola (Congo, 1976)). The difference between the range of the two methods is between 0 (Ebola (Bombali, 2014) and Ebola (Western Area Rural, 2014)), 0.2 (Ebola (Congco, 1976)). P parameter's range for LSQ method is between 0.13 (Ebola (Bomi, 2014)) and 0.98 (Ebola (Montserrado, 2014)); whereas, for the MLE method the range of p parameter is between 0.15 (Ebola (Bomi, 2014)) and 0.97 (Ebola (Bombali, 2014)). The difference between the range of the two methods is between 0 (Ebola (Bo, 2014), Ebola (Western Area, Rural, 2014)) and 0.18 (Ebola (Montserrado, 2014)).

Figure1-1 Parameter error bars. For each outbreak, the graphs show the mean and 95% confidential interval of r for LSQ and MLE methods.

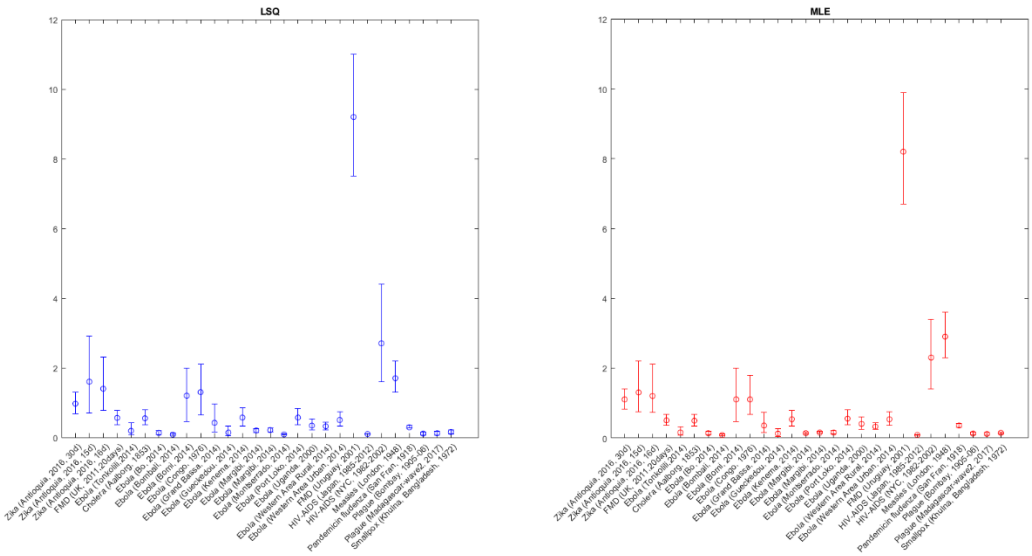
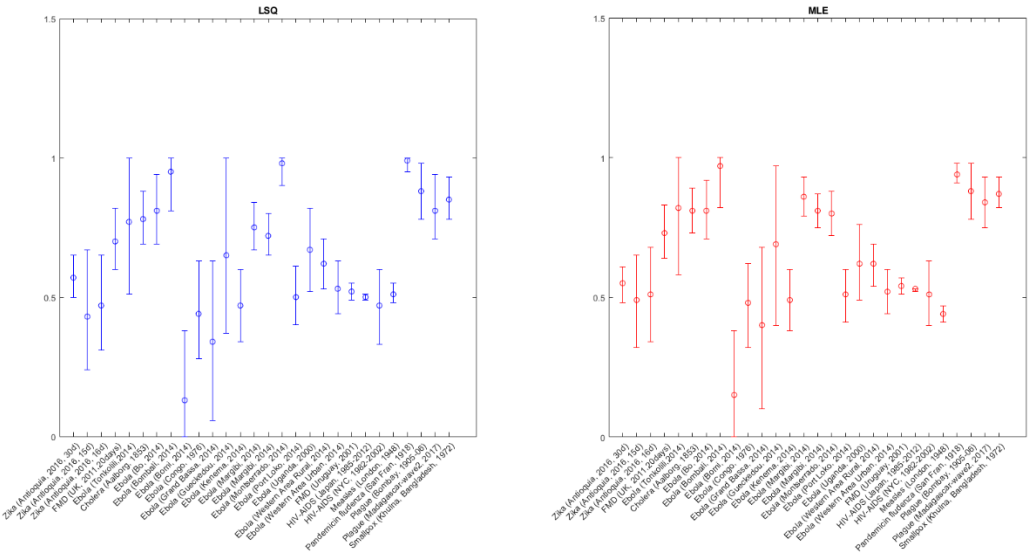


Figure1-2 Parameter error bars. For each outbreak, the graphs show the mean and 95% confidential interval of p for LSQ and MLE methods.



4. Discussion

Among the 28 outbreaks analyzed, the mean value for r shows similar results for the LSQ and MLE methods for most outbreaks (Figure 1-1). All but six outbreaks have higher mean estimates of r using the LSQ method in comparison to use the MLE method, but the differences are small (Table 3). A marked difference in the absolute value for r parameter estimates by the LSQ and MLE method is observed for the Measles outbreak (London, 1948), 1.2. Except 2 outbreaks, Measles (London, 1948) and FMD (Uruguay, 2001), the rest outbreaks show less than 1 for the r parameter difference between LSQ and MLE. With the result, both LSQ and MLE product the similar results for r parameter estimation.

20 of the total 28 outbreaks have higher estimated values of p with the LSQ method. However, the highest differing value of the p parameter is found to be 0.18 for the Ebola epidemic (Montserrado, 2014) with 0.98 for LSQ and 0.8 for MLE (Table 1,2, &3). This value is close to 0 but it is not a similar value regarding with the range of the p parameter, which was identified to be from 0 to 1. 26 of the outbreaks have an absolute mean difference less than 0.1 for the estimated parameter p value (Figure 1-2 & Table 3). Except 2 outbreaks, Ebola outbreak (Margibi, 2014, 10 weeks) and Ebola outbreaks (Montserrado, 2014), the rest outbreaks show less than 0.1 for the p parameter difference between LSQ and MLE. With the result, both LSQ and MLE product the similar results for p parameter estimation like r parameter result.

As mentioned at method part, RMSE and Anscombe residual indicate goodness-of-fit for the best fit model with the data with the estimated parameters. Of the total outbreaks, 15 show the value of RMSE to be higher with the LSQ method compared to the MLE method, though for all but 3 outbreaks, the absolute difference in RMSE values is observed to be lower than 10. LSQ has higher RMSE values for 15 outbreaks, indicating that MLE provided a better fit to the data. The absolute values of the RMSE come out to be less than 1 for half of the outbreaks, which concludes that LSQ and MLE methods perform comparably in terms of RMSE. The most

different result for RMSE between LSQ and MLE is estimated for the HIV-AIDS (Japan, 1985-2012). It ranges from 574.6561 for the LSQ method and 551.8599 for the MLE method, though this higher difference is expected with the high values of the RMSE for LSQ and MLE respectively. But the differences for the most outbreaks are small, indicating that the result of RMSE for two methods are similar.

For all but two outbreaks, MLE yielded higher values of Anscombe compared to LSQ. The highest difference is observed for the HIV-AIDS epidemic (Japan, 1985-2012), 59.2761. Most outbreaks show a difference of less than 10 for the Anscombe residual when estimated using LSQ versus MLE methods. Like RMSE, Anscombe residual indicates that lower value shows a better fit to the data. With the indication, for most outbreaks, using MLE provides better fit to the data. But the differences between two methods are small for most outbreaks, indicating that Anscombe residuals are similar with two methods.

Even Anscombe residual would be better for the Poisson distribution, both RMSE and Anscombe residual assess the goodness of fit, so the result tendency would show similar. For example, the outbreak that has the most different value is HIV-AIDS epidemic (Japan, 1985-2012). However, this result also indicates that GGM would not a good model for this outbreak since the HIV at Japan outbreak and Measles outbreaks at London have very high values for both variables respectively.

For the 95% prediction interval coverage, most outbreaks show similar results between LSQ and MLE. The prediction interval coverage shows how the GGM fit with the data. Even though the difference in the prediction interval estimates using LSQ and MLE methods for Ebola (Monteserrado, 2014) is smaller than that for the Pandemic Influenza (San Francisco, 1918), the number of data point terms seems different. The total of data points used for the analysis of Ebola (Montserrado, 2014) is 10 and those used for the analysis of Pandemic Influenza (San Francisco, 1918) were 20. Therefore, the prediction coverage percentage for the Pandemic Influenza (San Francisco, 1918) is estimated to be lower than for the Ebola (Montserrado, 2014) epidemic.

For some outbreaks, the 95% prediction interval coverages are low; however, the different result show that there is not different since 95% prediction interval coverage for both methods are same. For example, HIV-ADIS (Japan, 1985-2012) has 18.18% for LSQ and MLE methods. Even though the coverages are low for each method, the result shows that there is no different. In terms of the RMSE and Anscombe residual, some outbreaks have high values; for example, HIV-ADIS (Japan, 1985-2012) has around 550 RMSE and around 700 Anscombe residual for both methods, which mean the model is not fit with the data.

The result would not adapt to the other model such as generalized logistic growth model (LGM) since this study is used for GGM and estimate just r and p parameters. Further, in terms of the ascending phase, two outbreaks, Zika (Antioquia, 2016) and Ebola (Margibi, 2014) are used with different data points. Just Zika (Antioquia, 2016) with 30 days from with 15 and 16 days has different results of r and p parameters, RMSE, Anscombe residual and prediction interval coverage with LSQ method and p parameter, RMSE, and Anscombe residual with MLE method. But it is hard to say that the ascending phase affect the results of the methods comparably since it

was one outbreak. Future studies could include comparing the parameter estimation with other models and the effect of the ascending phase into the results.

Overall, results demonstrate that utilizing LSQ and MLE methods deduct similar results for the GGM parameters r and p, RMSE, Anscombe residual, and prediction coverage for most outbreaks despite showing a large difference for a few. It is also demonstrated that the number of data points effect the results; however, we do not have enough information on how to calculate the exact number of data points required.

References

(1) R. Anderson and R. May. *Infectious Diseases of Humans: Dynamics and Control*. New York; Oxford University Press, 1991.

(2) Xu, J., Murphy, S. L., Kochanek, K. D., Bastian, B., and Arias, E. *Deaths: Final Data for 2016*. 2018.

(3) G. Chowell, C. Viboud, J. M. Hyman, and L. Simonsen. *The western Africa ebola virus disease epidemic exhibits both global exponential and local polynomial growth rate*. 2014.

(4) J. Poorolajal, E. Hooshmand, H. Mahjub, N. Esmailnasab and E. Jenabi. *Survival rate of AIDS disease and mortality in HIV-infected patients: a meta-analysis.* 2016.

(5) K. Roosa, R. Luo, and G. Chowell. *Comparison of parameter estimation methods in the presence of overdispersion: a simulation study*. 2019

(6) C. Viboud, L. Simonsen and G. Chowell. *A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks* 2016

(7) A. Gordon, L. Gresh, S. Ojeda, L.C. Katzlnick, N. Sanchez, Juan Carlos Mercado, G. Chowell, B. Lopez, D. Elizondo, J. Coloma, R. Burger-Calderon, G. Kuan, A Balmaseda, and E. Harris. *Prior dengue virus infection and risk of Zika: A pediatric cohort in Nicaragua*. 2019

(8) D.W. Shanafelt, G. Jones, M. Lima, C. Perrings, and G. Chowell *Forecasting the 2001 Foot-and-Mouth Disease Epidemic in the UK*. 2018

(9) G. Chowell, C. Viboud, J.M. Hyman and L. Simonsen. *The Western Africa ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates*. 2015

(10) G. Chowell, D. Hincapie-Palacio, J. Ospina, B. Pell, A. Tariq, S. Dahal, S. Moghadas, A. Smirnova, L. Simonsen, and C. Viboud. *Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics* 2016

(11) L. Dinh, G. Chowell, and R. Rothernberg. *Growth scaling for the early dynamics of HIV/AIDS epidemics in Brazil and the influence of socio-demographic factors* 2018

(12) Gerado Chowell. *Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability and forecasts* 2017

(13) In Jae Myung. *Tutorial on maximum likelihood estimation* 2002

(14) Konstantin Kashin. *Statistical inference: Maximum likelihood estimation* 2014

(15) Max Kuhn and Kjell Johnson. *Applied predictive modeling* 2013

(16) P. McCullagh and J.A. Melder. *Generalized linear models* 1983

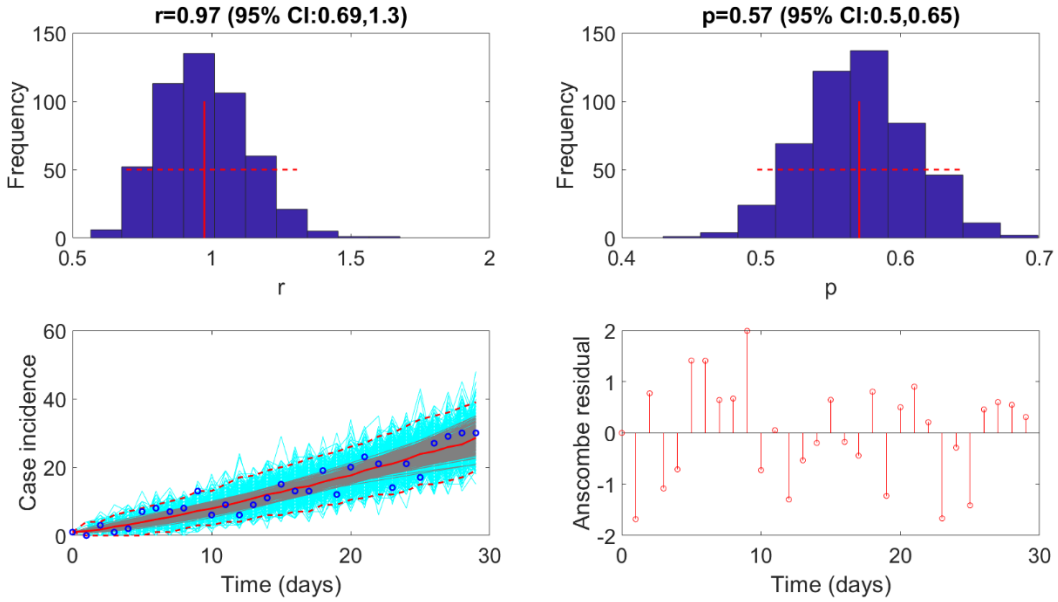
(17) Kimberlyn Roosa and Gerardo Chowell. *Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models*. Theoretical Biology and Medical Modelling, 16(1):1, 2019.

APPENDIX

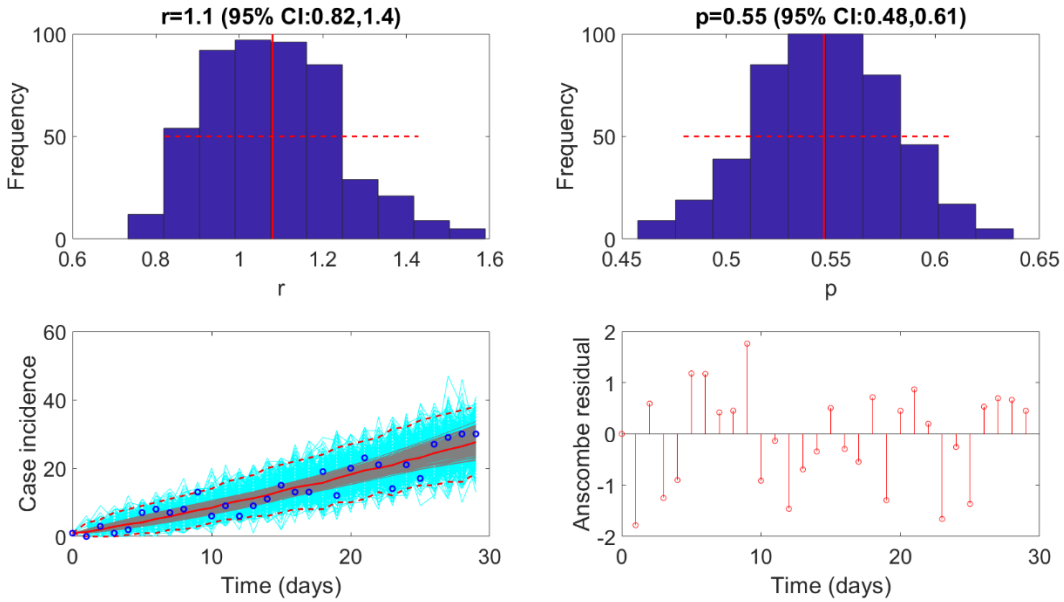
APPENDIX

FigrueS1-1. GGM parameter estimates as each outbreak. Parameters r and p estimates and 95% confidence intervals are represented at upper two graphs. For the r and p parameters graphs, the red line means mean of the results and red dash as the 95% CI of the results. Left below figure shows the fitting GGM model with blue bubbles as the data, red dash line as 95% prediction interval, red line as mean of bootstrap, grey lines as bootstraps, and cyan as prediction intervals bootstraps

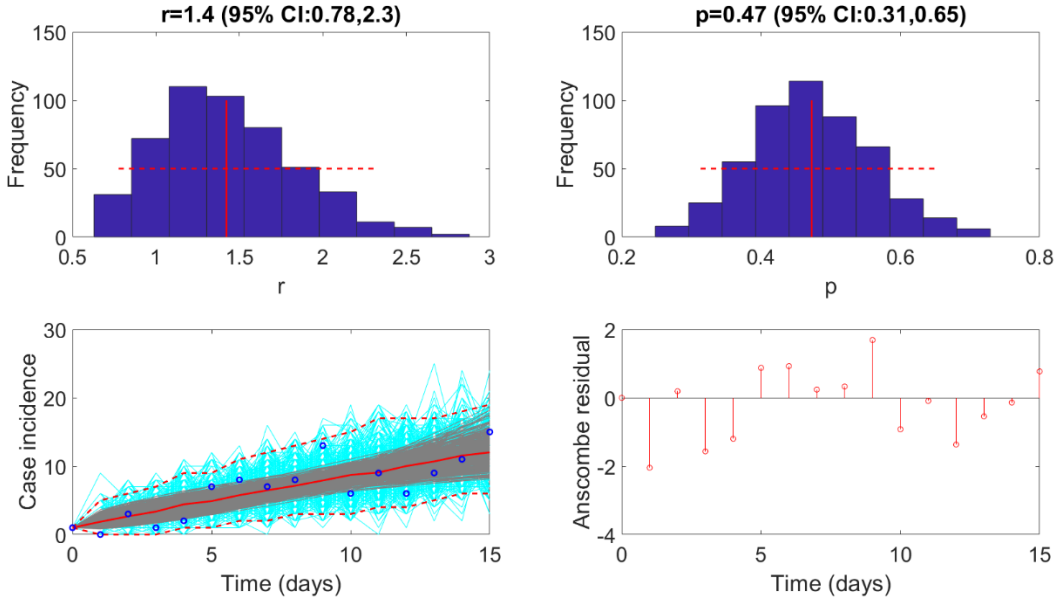
a-1. Zika (Antioquia, 2016, (30d) with LSQ method



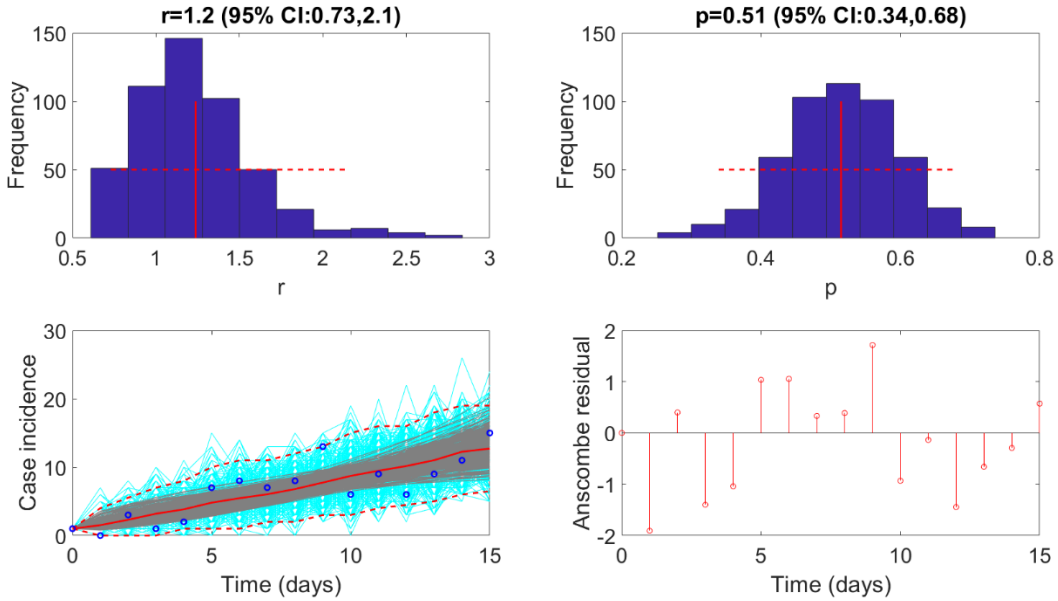
a-2. Zika (Antioquia, 2016, 30d) with MLE method



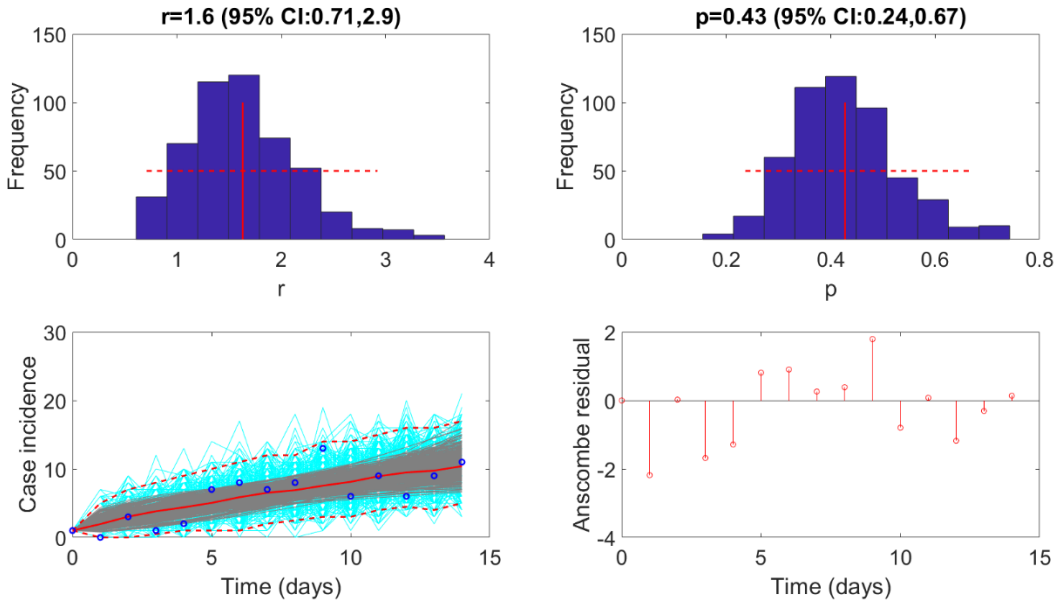
b-1 Zika (Antioquia, 2016, 16d) with LSQ method



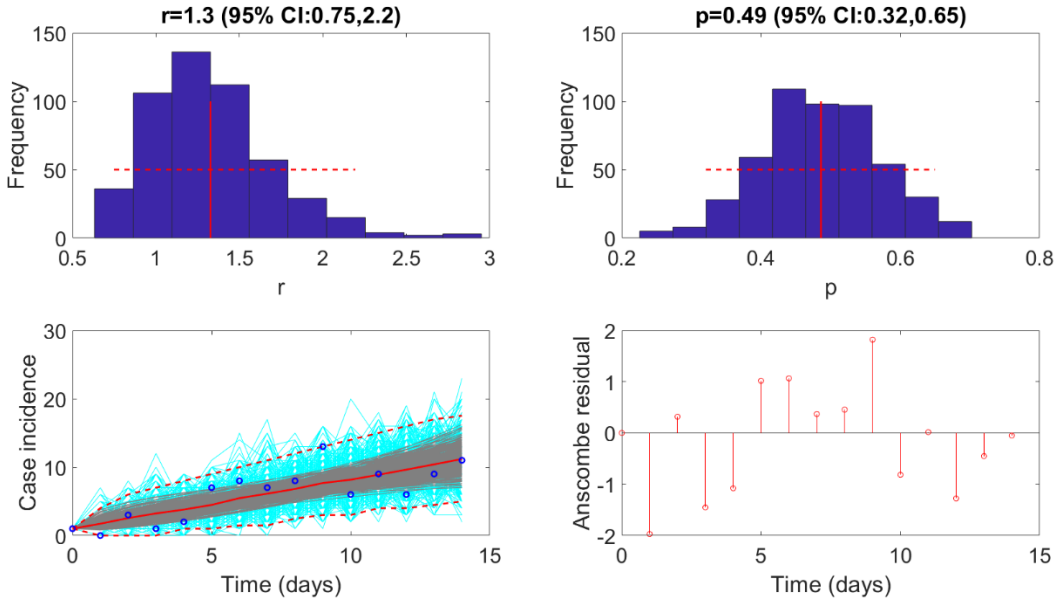
b-2 Zika (Antioquia, 2016, 16d) with MLE method



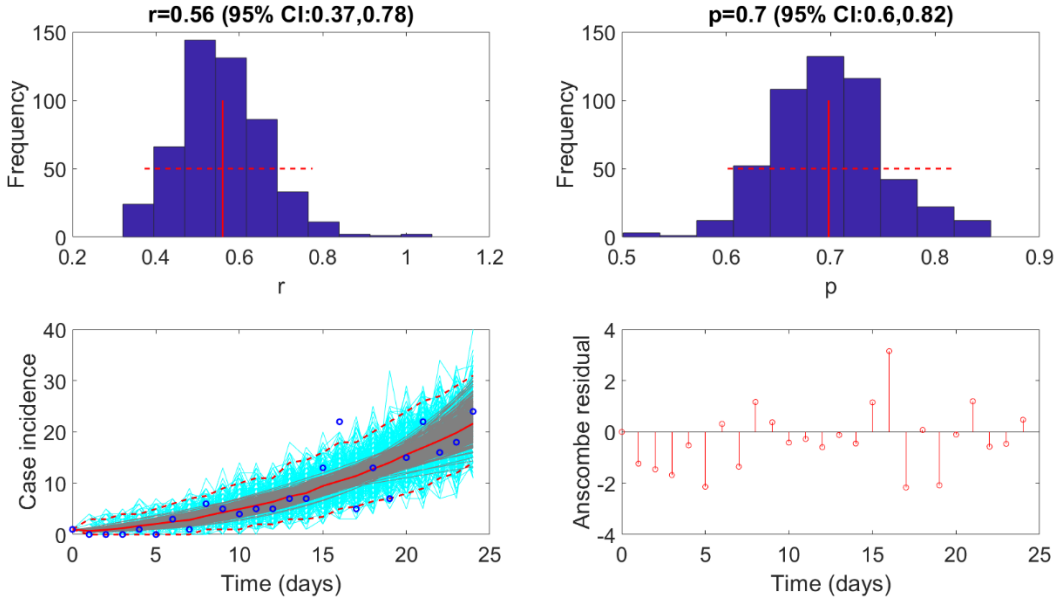
c-1 Zika (Antioquia, 2016, 15d) with LSQ method



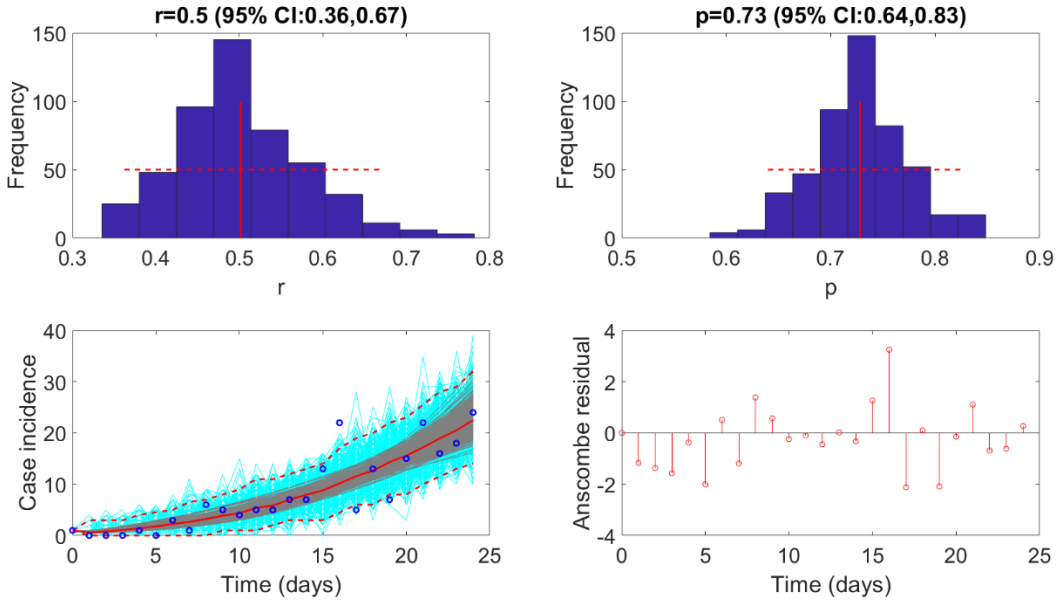
c-2 Zika (Antioquia, 2016, 15d) with MLE method



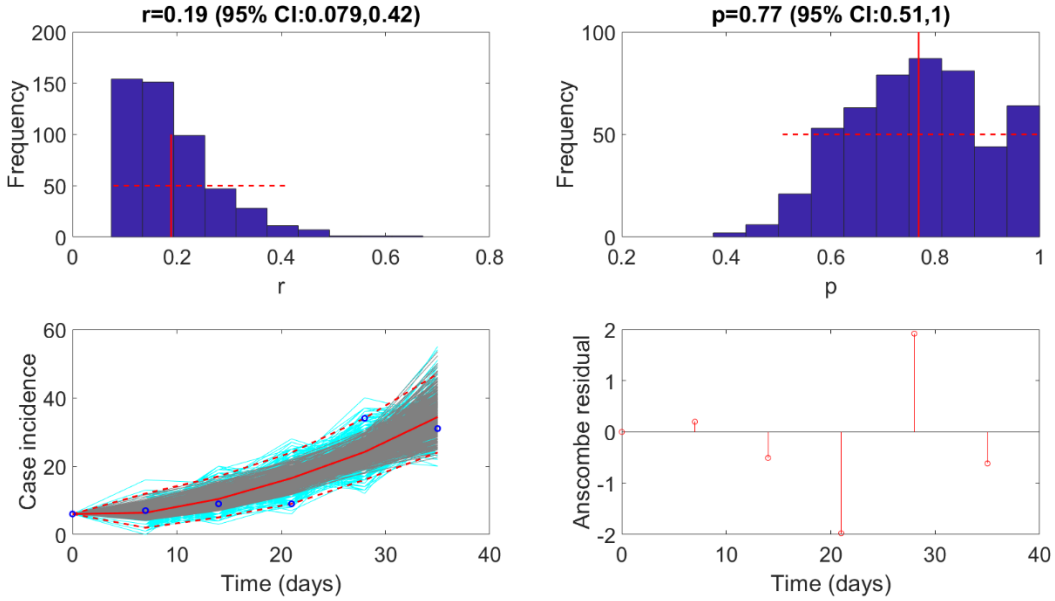
d-1 FMD (UK,2001-120days, 25d) with LSQ method



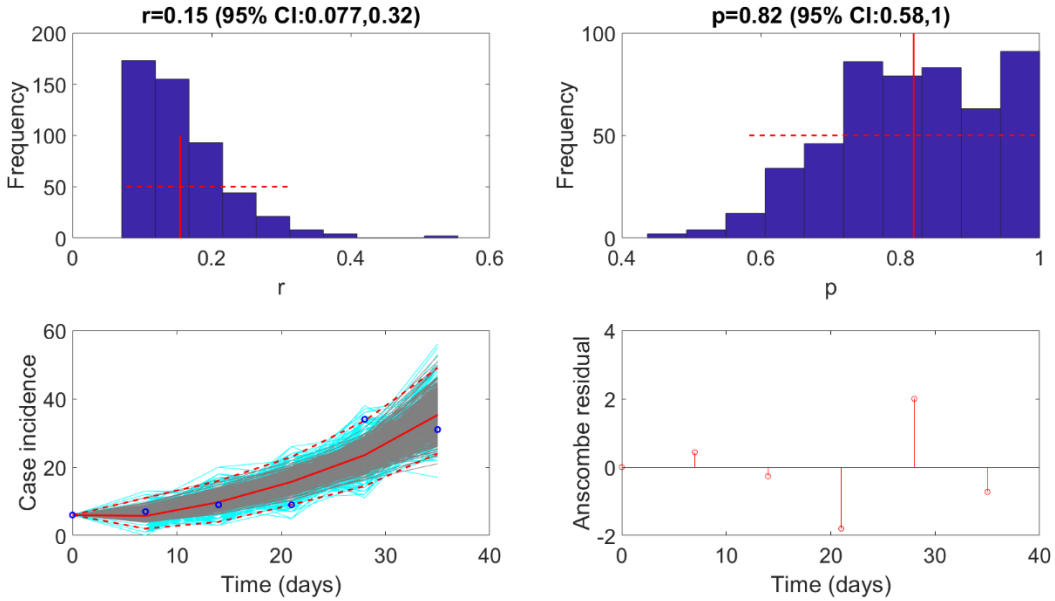
d-2 FMD (UK,2001-120days, 25d) with MLE method



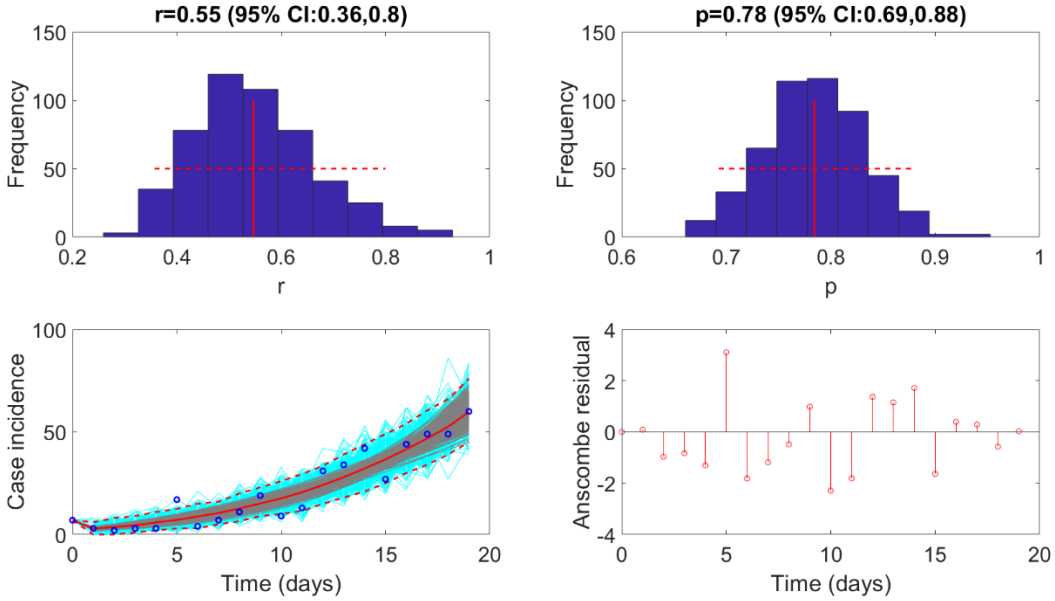
e-1 Ebola (Tonkolili, 2014, 6w) with LSQ method



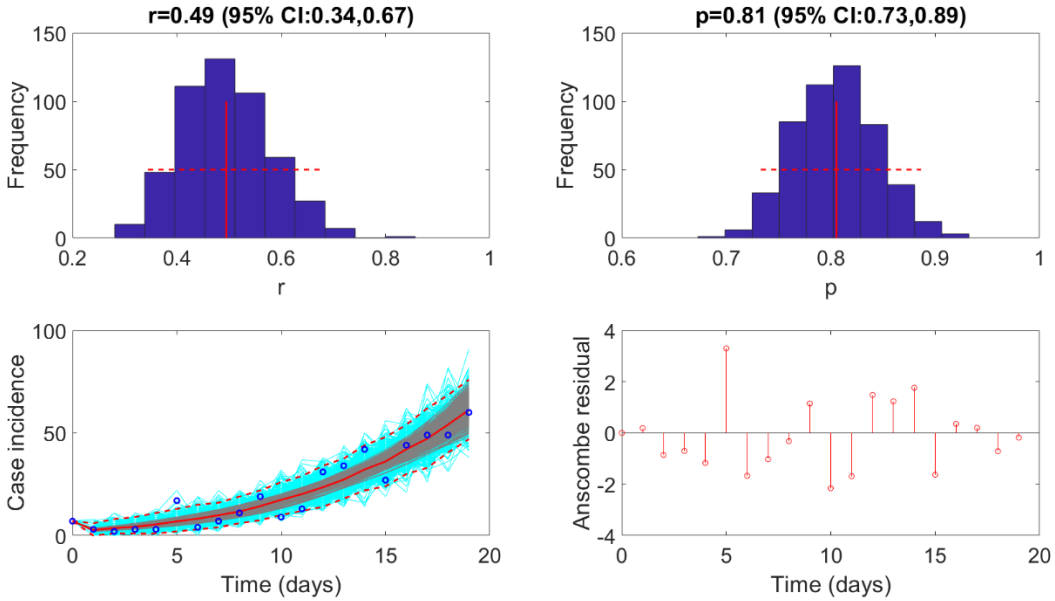
e-2 Ebola (Tonkolili, 2014, 6w) with MLE method



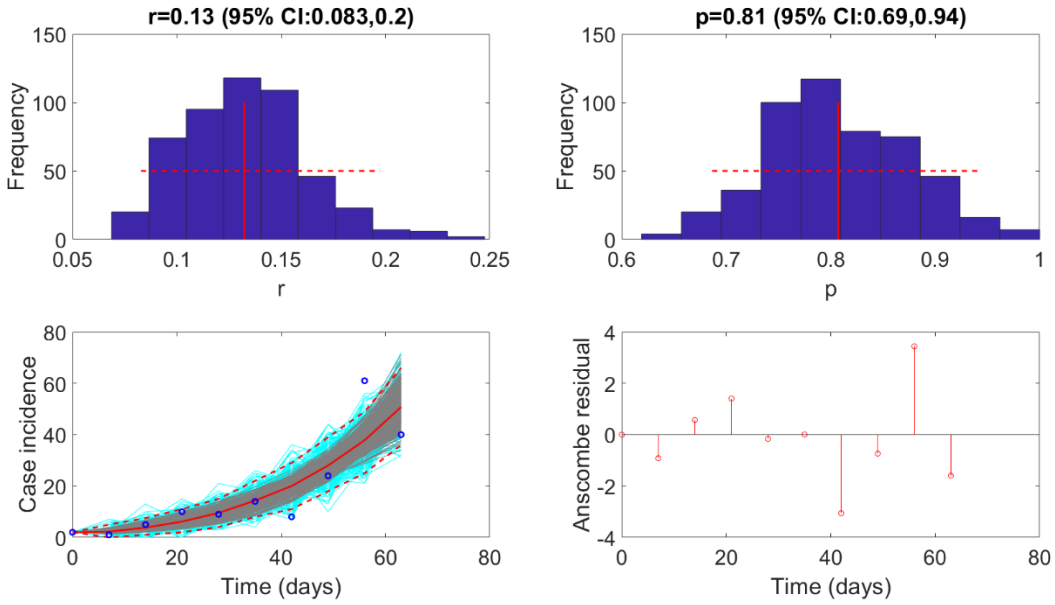
f-1 Cholera (Aalborg,1853, 20d) with LSQ method



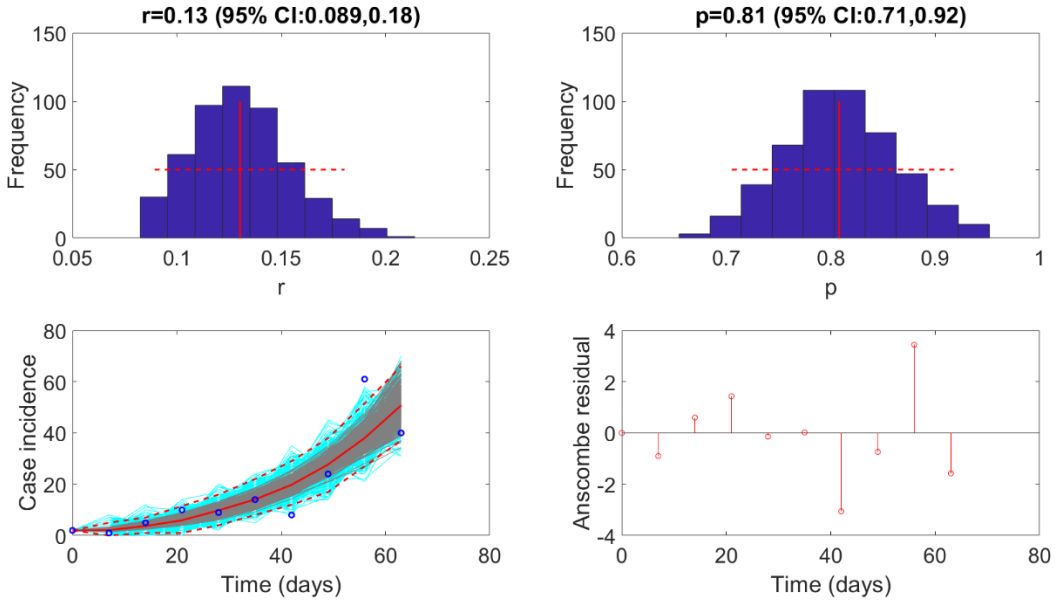
f-2 Cholera (Aalborg, 1853, 20d) with MLE method



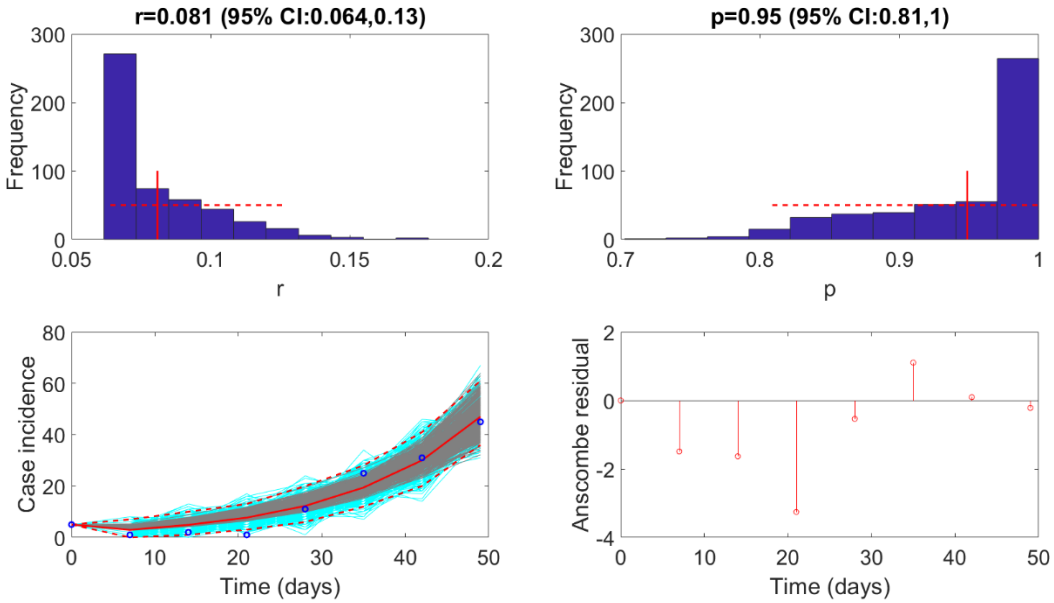
g-1 Ebola (Bo, 2014) with LSQ method



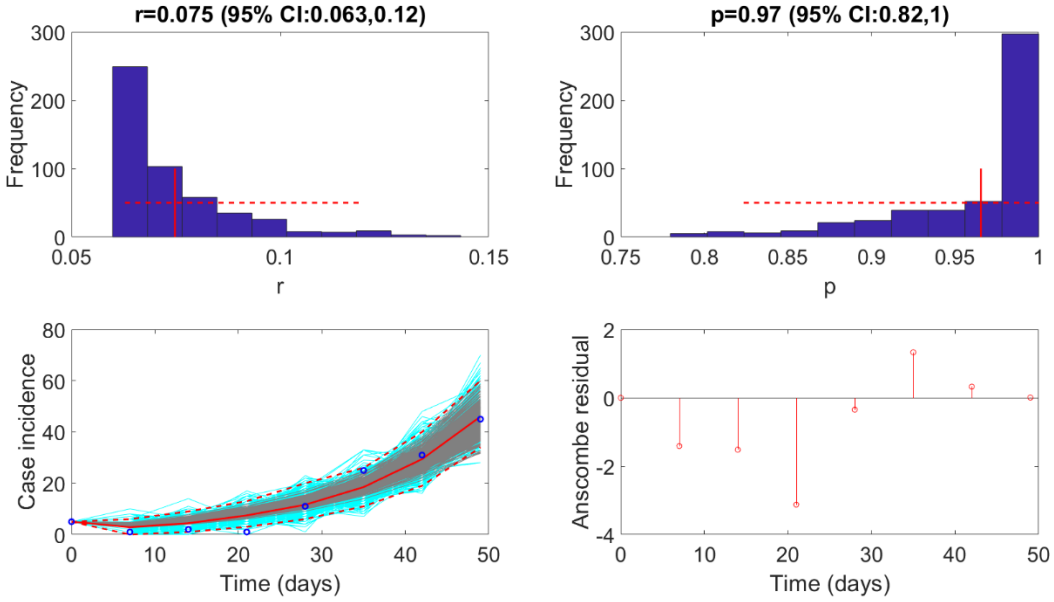
g-2 Ebola (Bo, 2014) with MLE method



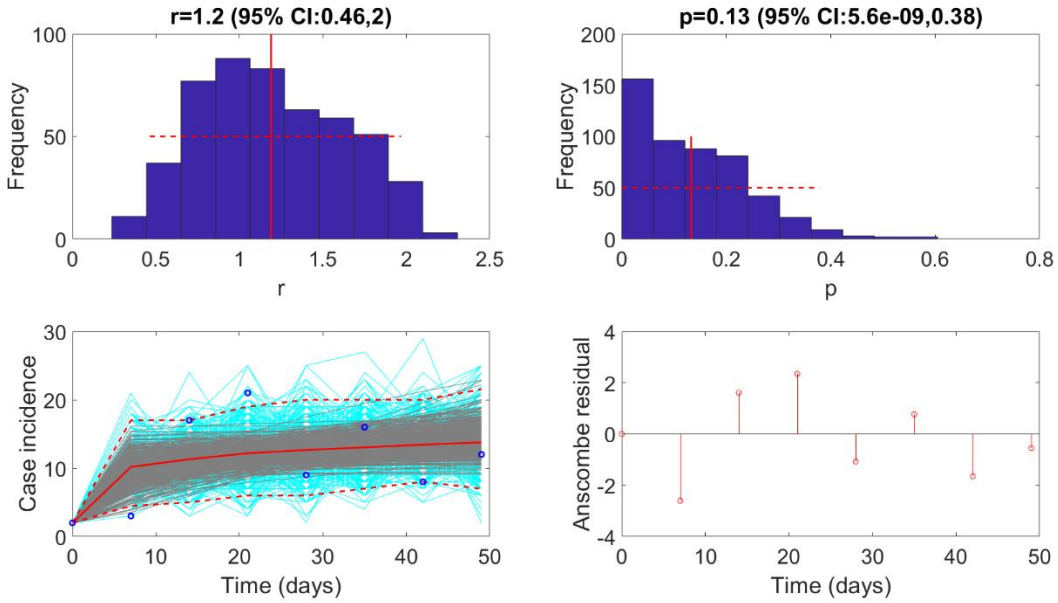
h-1 Ebola (Bombali, 2014) (8w) with LSQ method



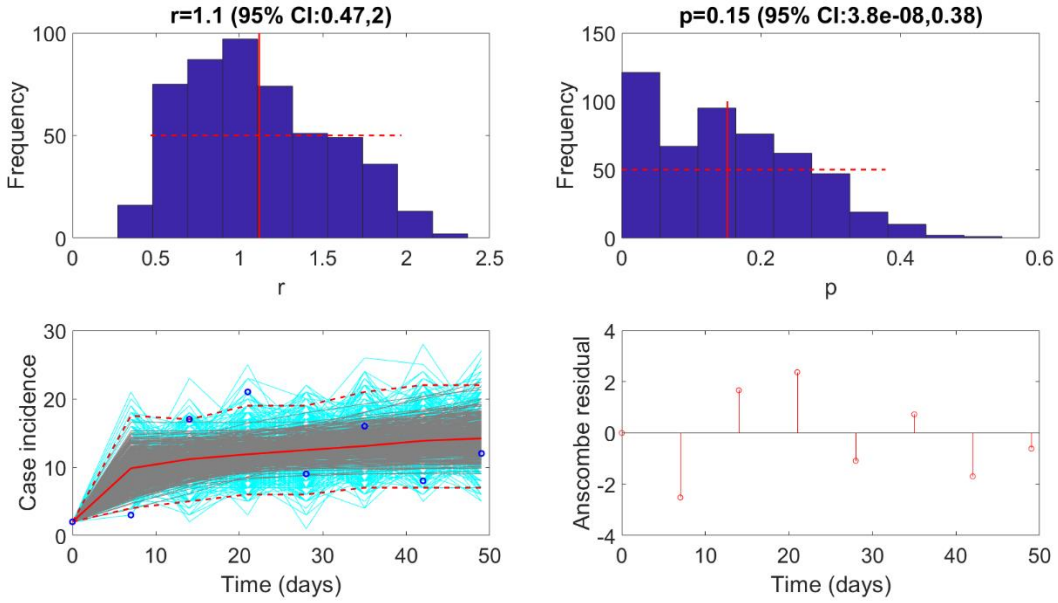
h-1 Ebola (Bombali, 2014) (8w) with MLE method



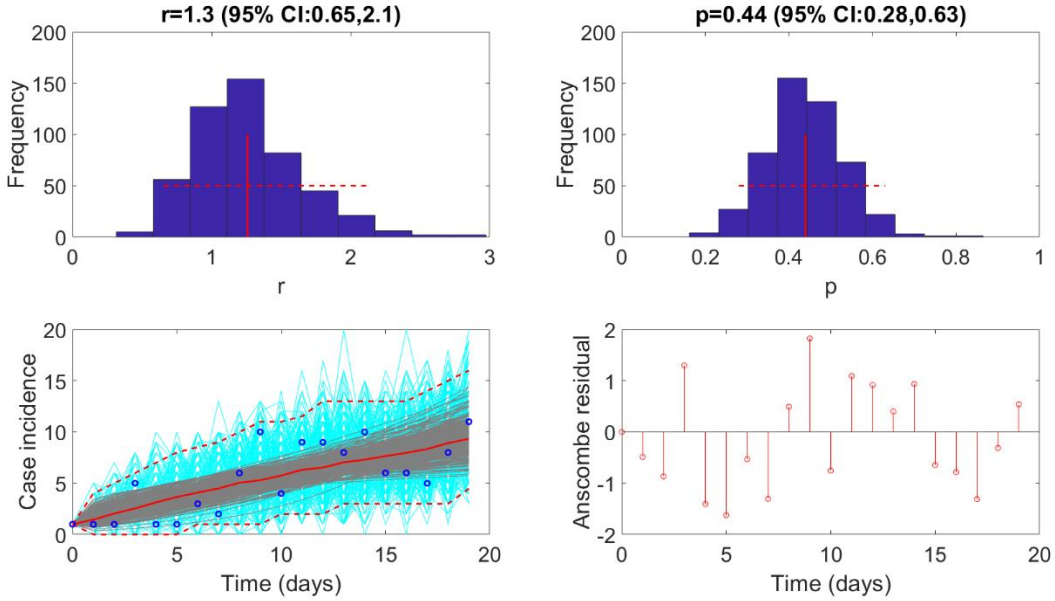
i-1 Ebola (Bomi, 2014) (8w) with LSQ method



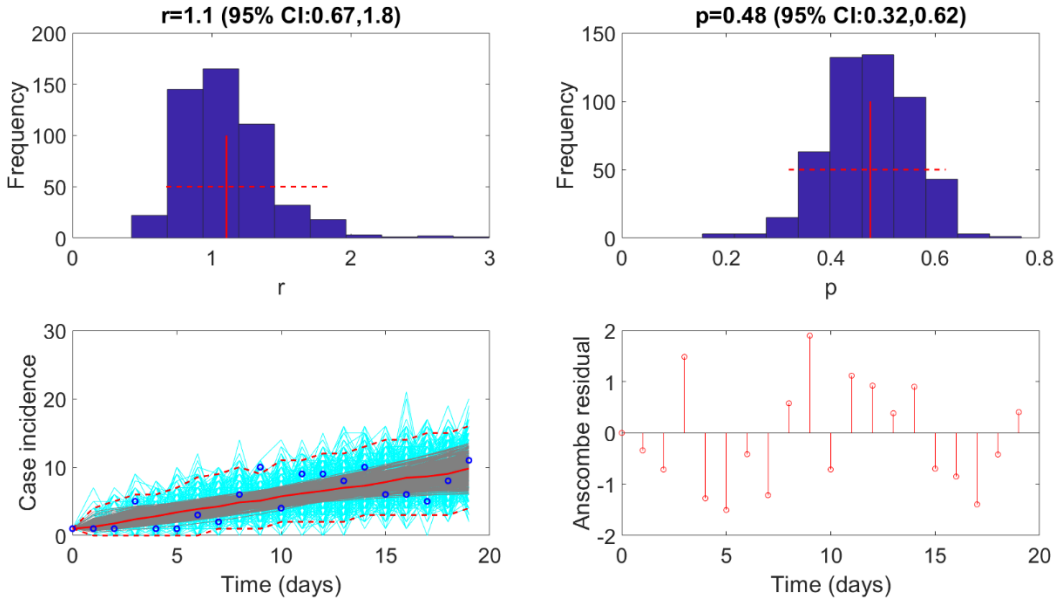
i- 2 Ebola (Bomi, 2014) (8w) with MLE method



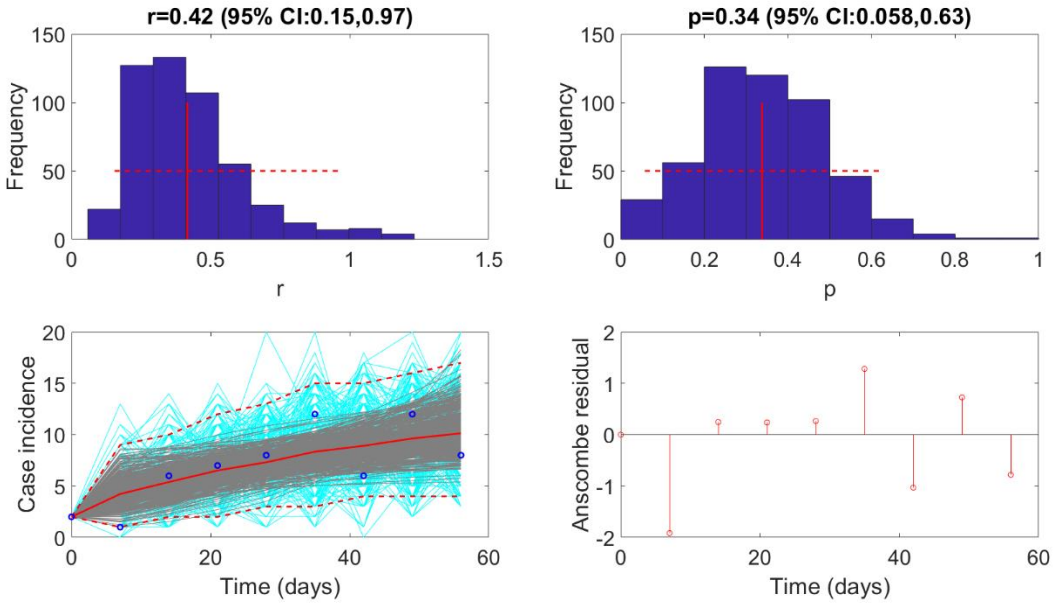
j-1 Ebola (Congo, 1976) (20d) with LSQ method



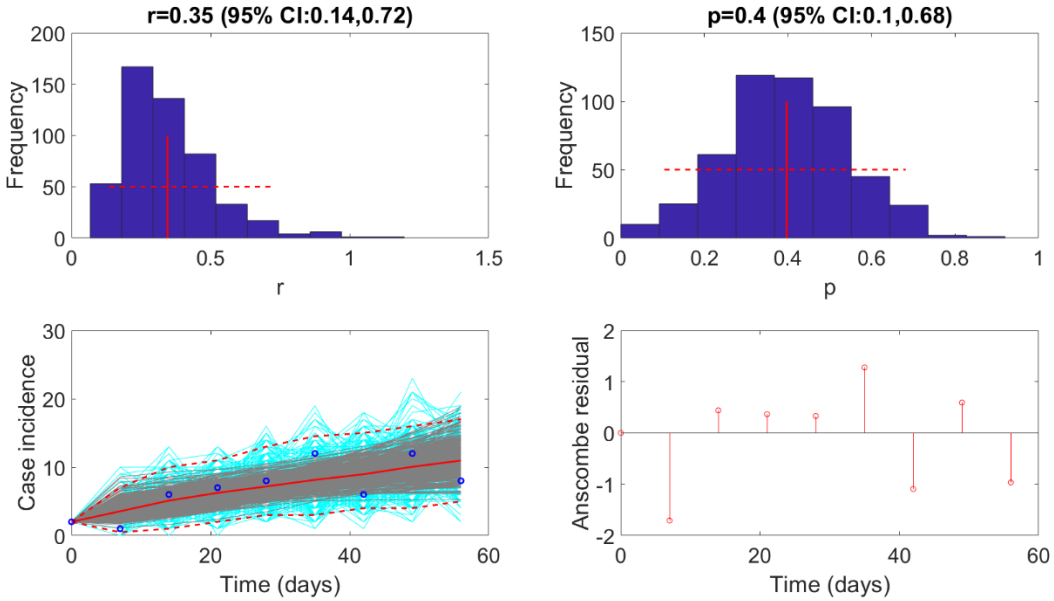
j-2 Ebola (Congo, 1976) (20d) with MLE method



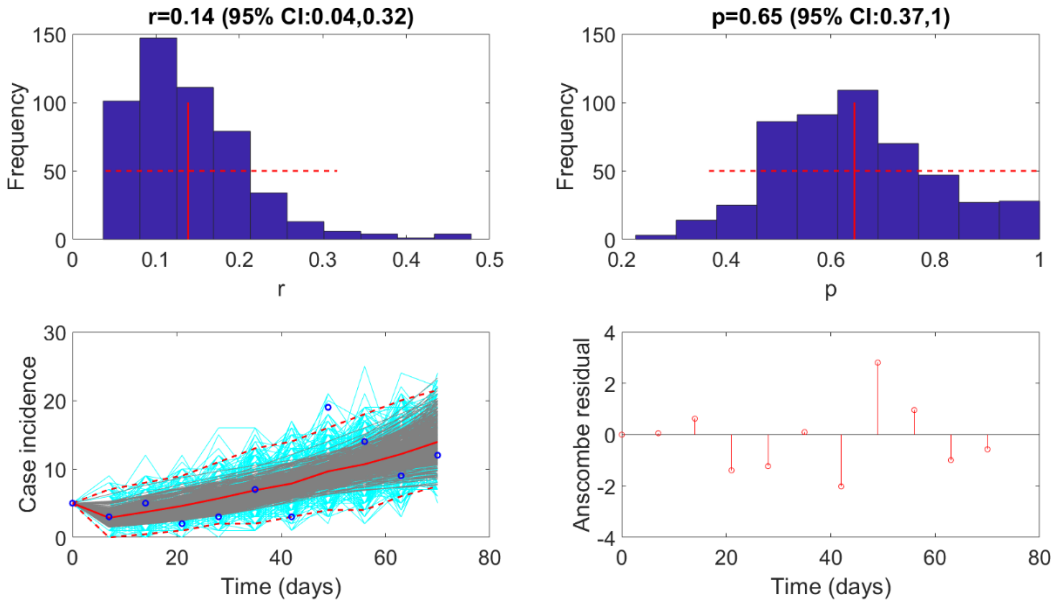
k-1 Ebola (Grand Bassa, 2014) wih LSQ method



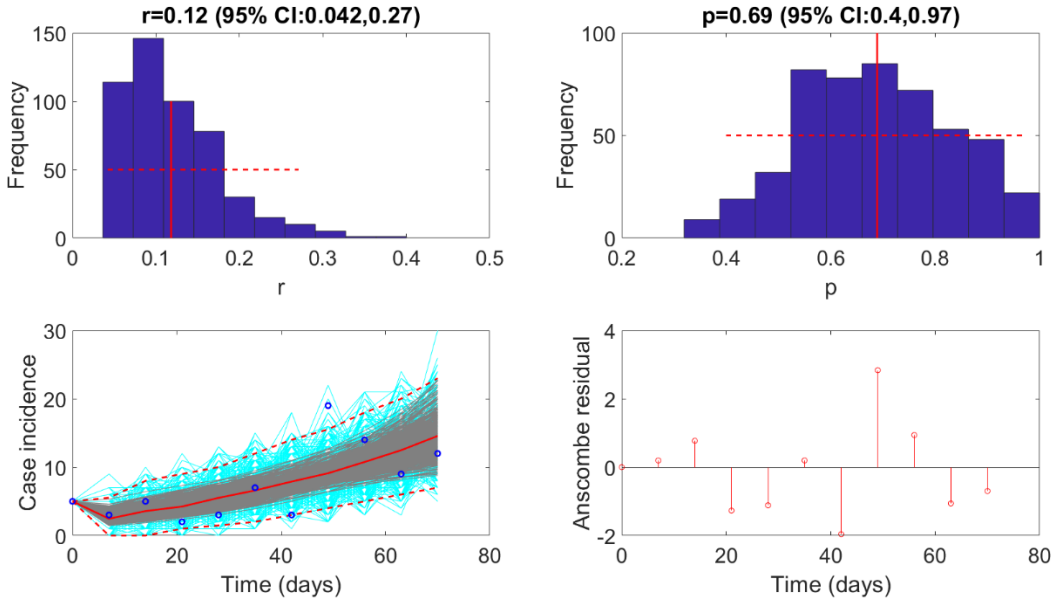
k-2 Ebola (Grand Bassa, 2014) wih MLE method



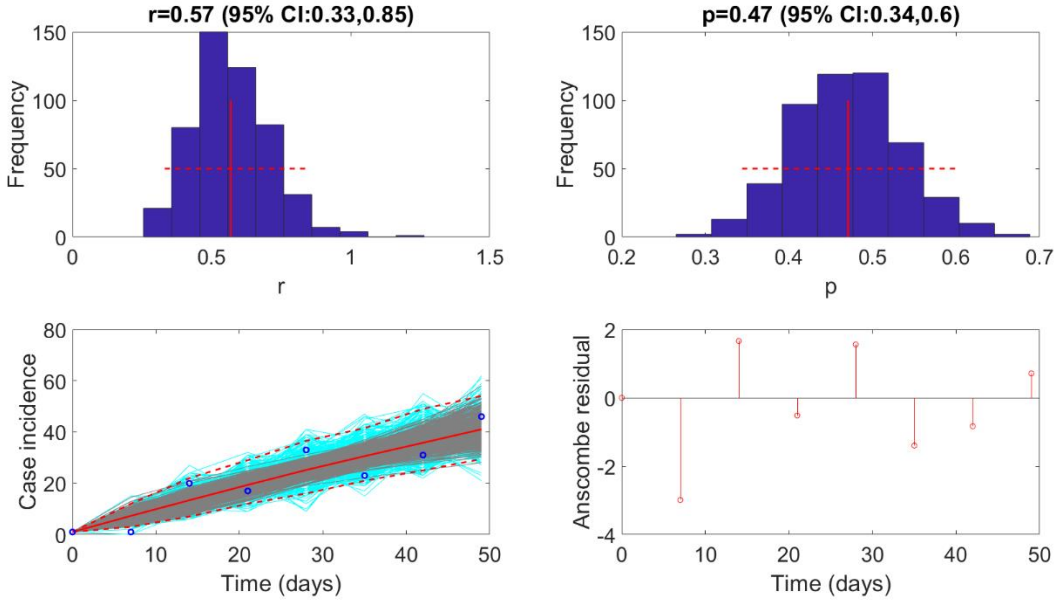
l-1 Ebola (Gueckedou, 2014) (11w) with LSQ method



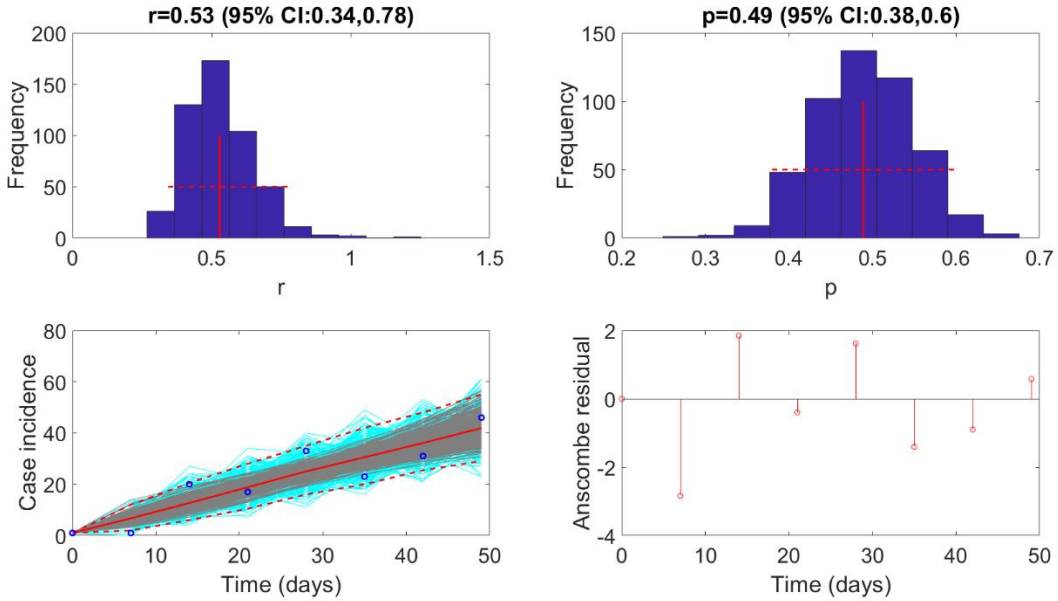
l-2 Ebola (Gueckedou, 2014) (11w) with MLE method



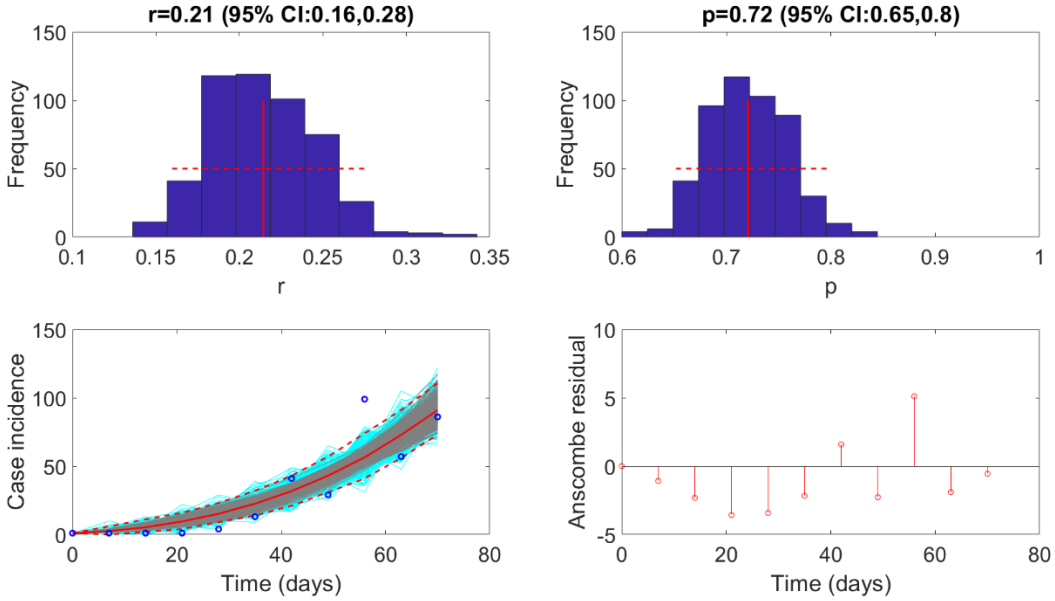
m-1 Ebola (Kenema, 2014) (8w) with LSQ method



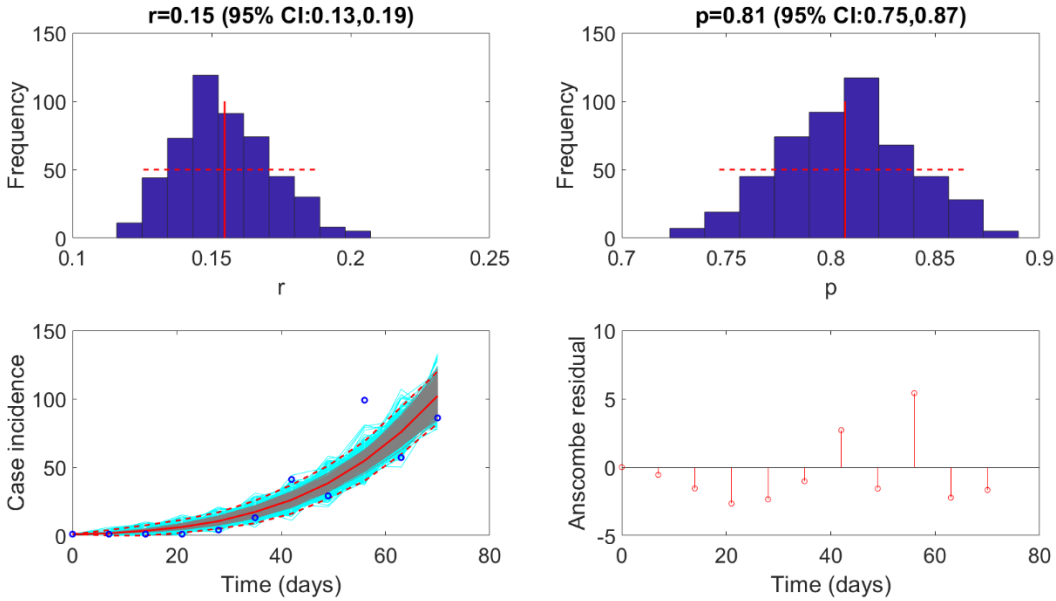
m-2 Ebola (Kenema, 2014) (8w) with MLE method



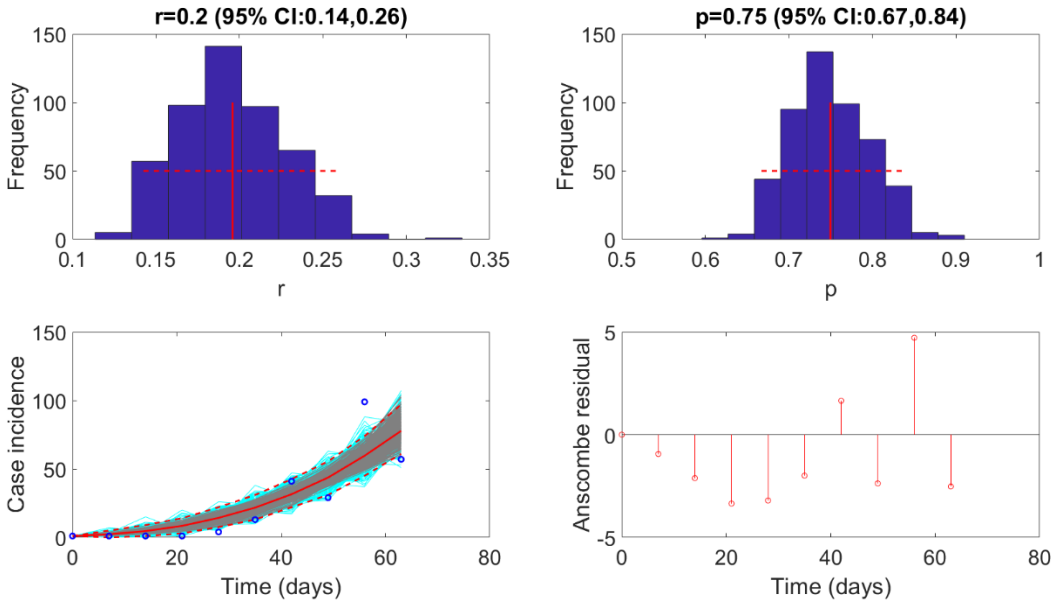
n-1 Ebola (Margibi, 2014) (11w) with LSQ method



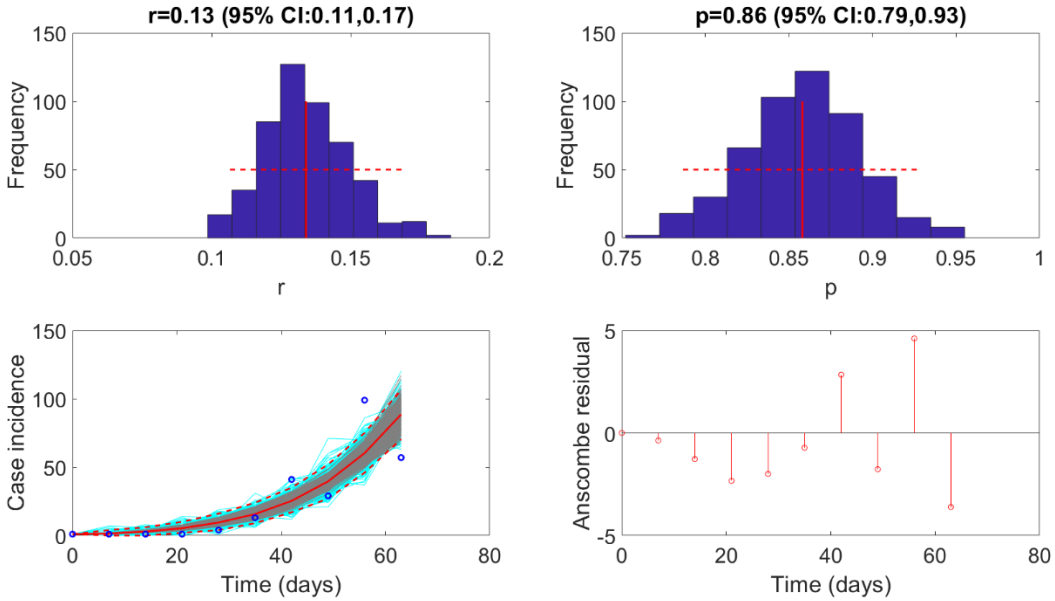
n-2 Ebola (Margibi, 2014) (11w) with MLE method



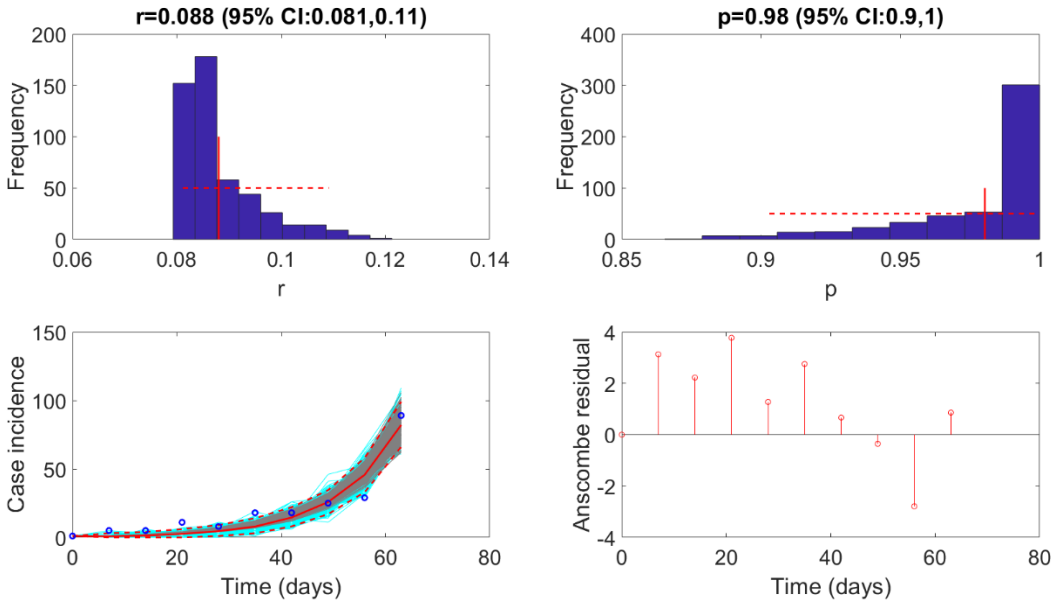
o-1 Ebola (Margibi, 2014) (10w) with LSQ method



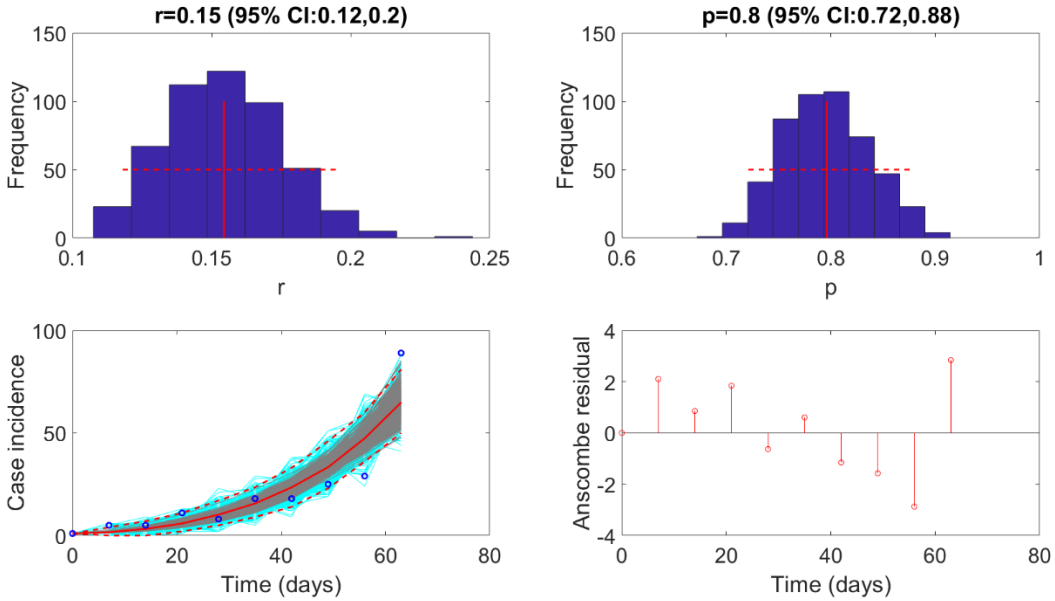
o-2 Ebola (Margibi, 2014) (10w) with MLE method



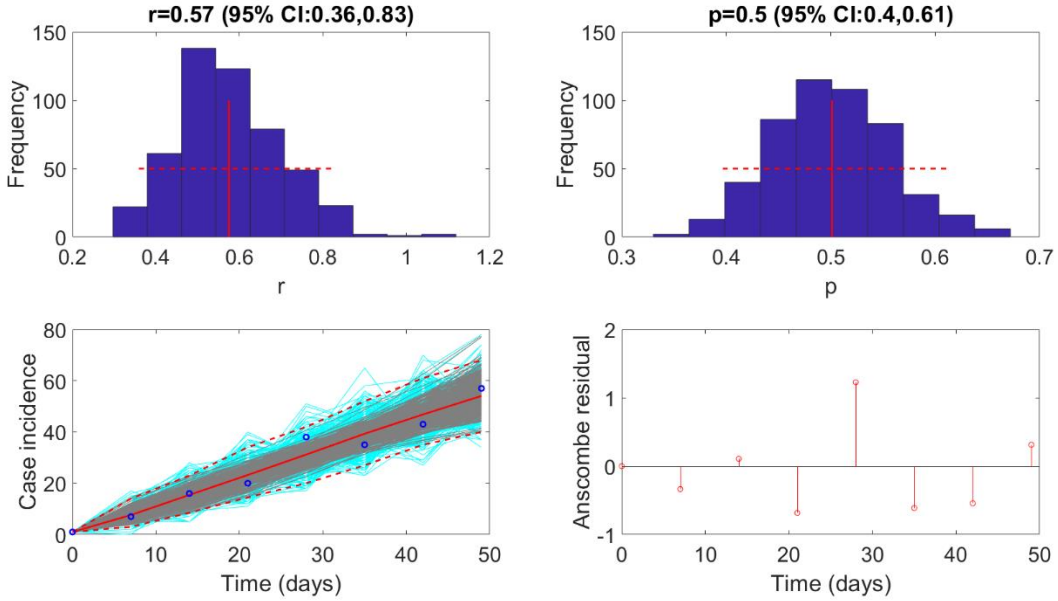
p-1 Ebola (Montserrado, 2014) with LSQ method



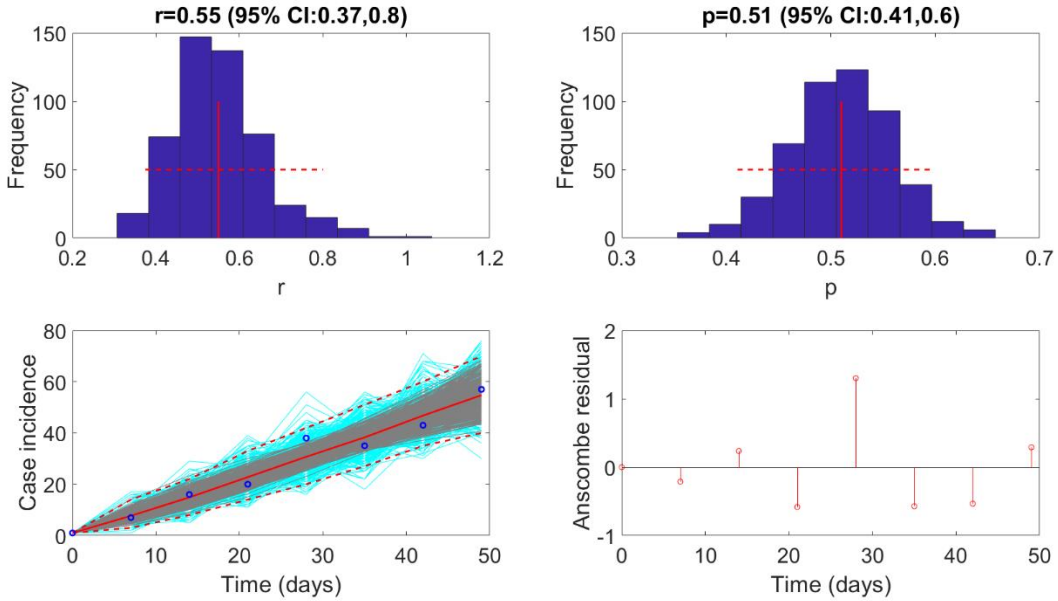
p-2 Ebola (Montserrado, 2014) with MLE method



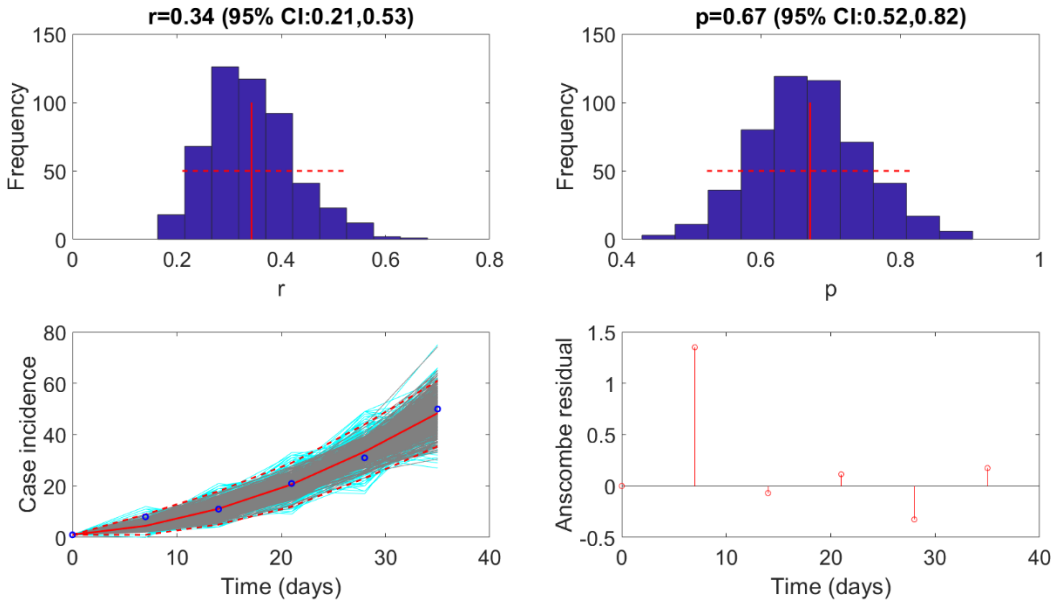
q-1 Ebola (Port Loko, 2014) (8w) with LSQ method



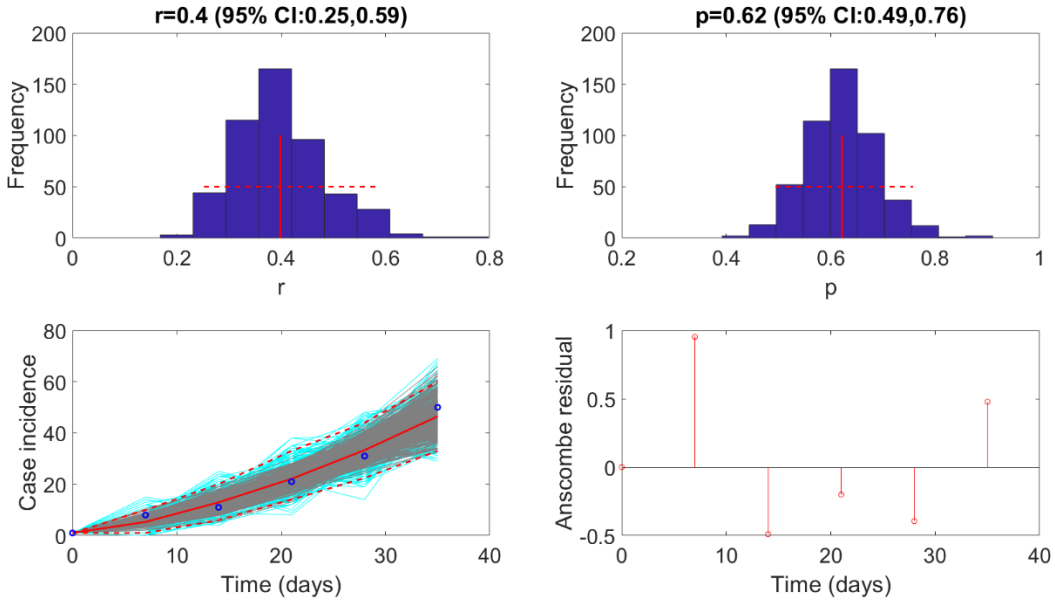
q-2 Ebola (Port Loko, 2014) (8w) with MLE method



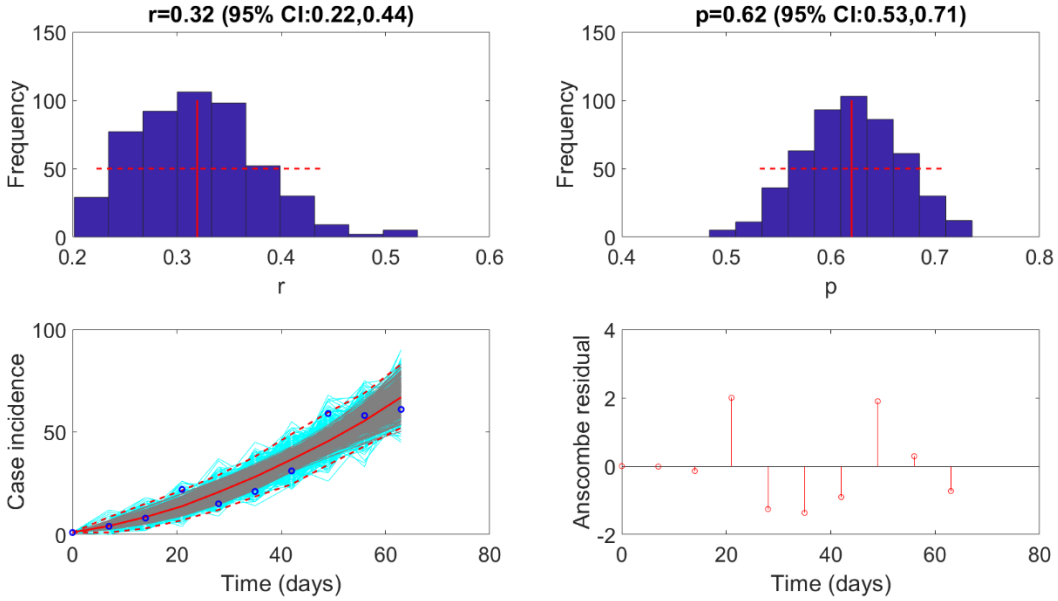
r-1 Ebola (Uganda, 2000) (6w) with LSQ method



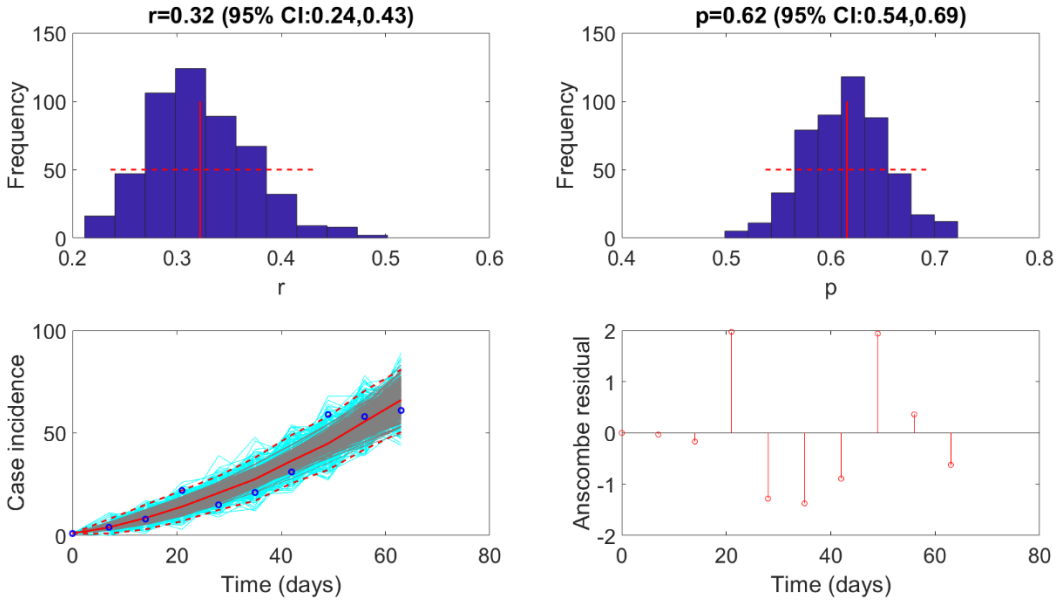
r-2 Ebola (Uganda, 2000) (6w) with MLE method



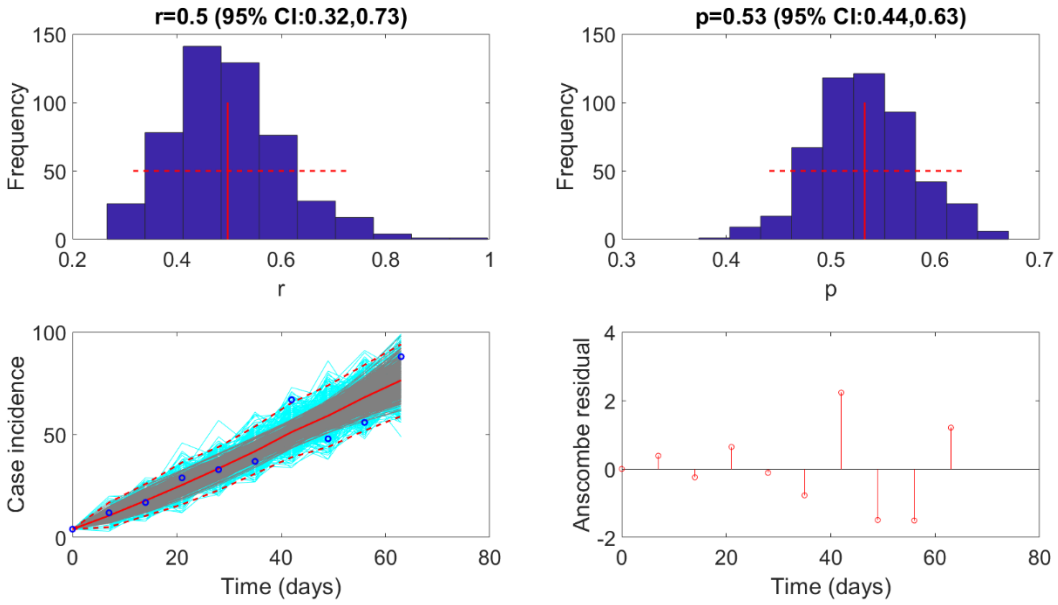
s-1 Ebola (Western Area Rural, 2014) (10w) with LSQ method



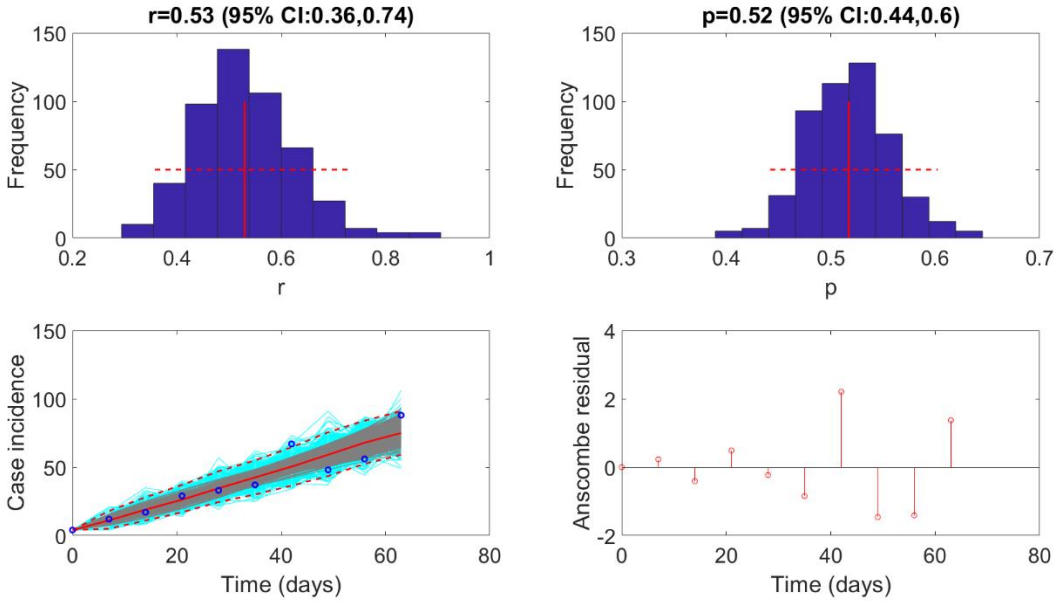
s-2 Ebola (Western Area Rural, 2014) (10w) with MLE method



t-1 Ebola (Western Area Urban, 2014) (10w) with LSQ method

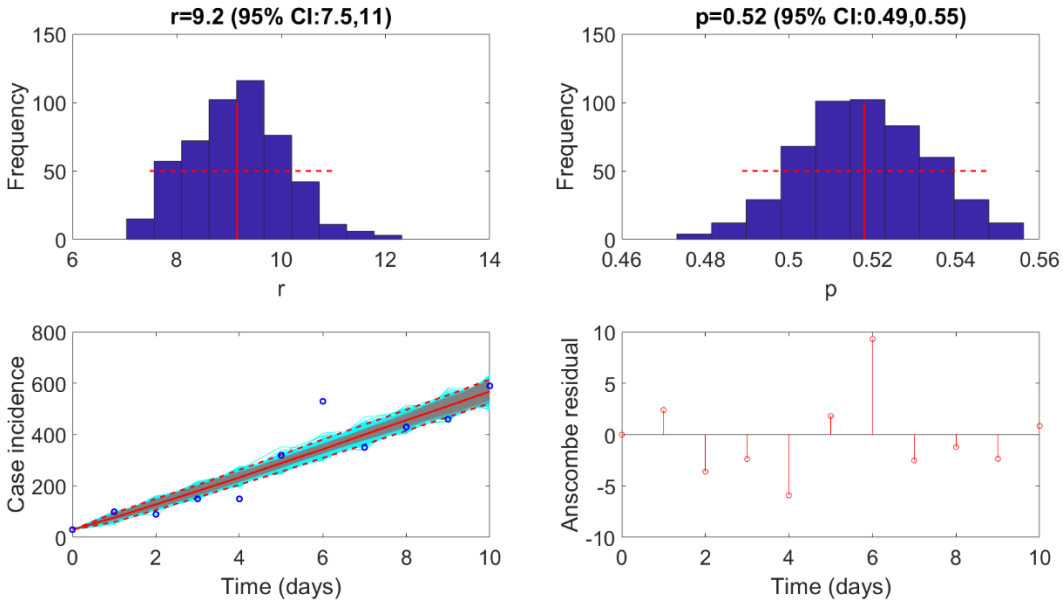


t-1 Ebola (Western Area Urban, 2014) (10w) with MLE method

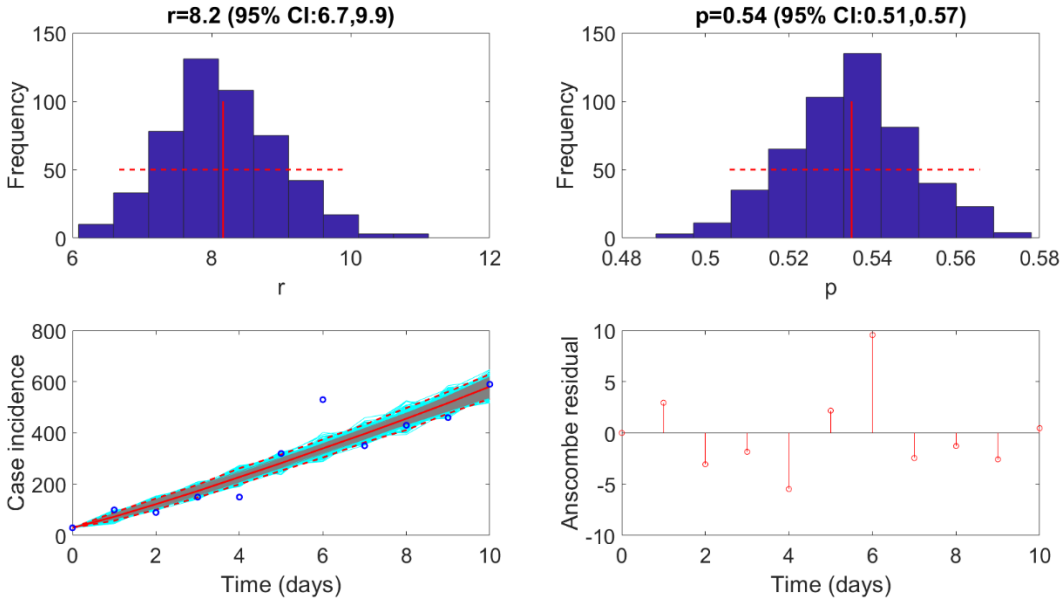


Figrue1-2. GGM parameter estimates as each outbreaks (cont.) Parameters r and p estimates and 95% confidence intervals are represented at upper two graphs and left below figure shows the fitting GGM model with blue bubbles as the data, red dash line as 95% prediction interval, red line as mean of bootstrap, grey lines as bootstraps, and cyan as prediction intervals bootstraps

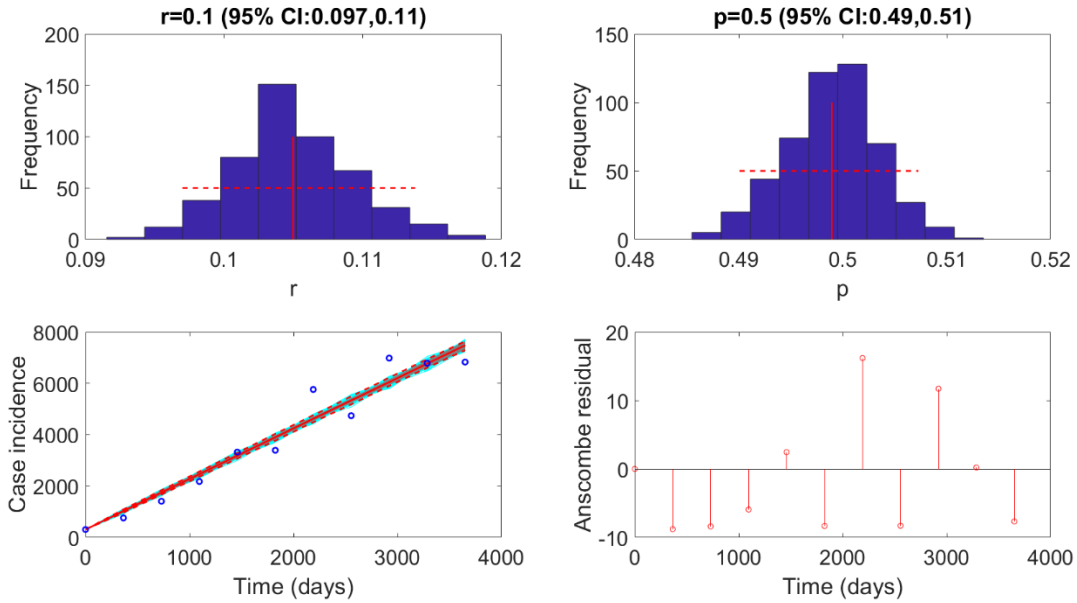
a-1 FMD (Uruguay, 2001) (10w) with LSQ method



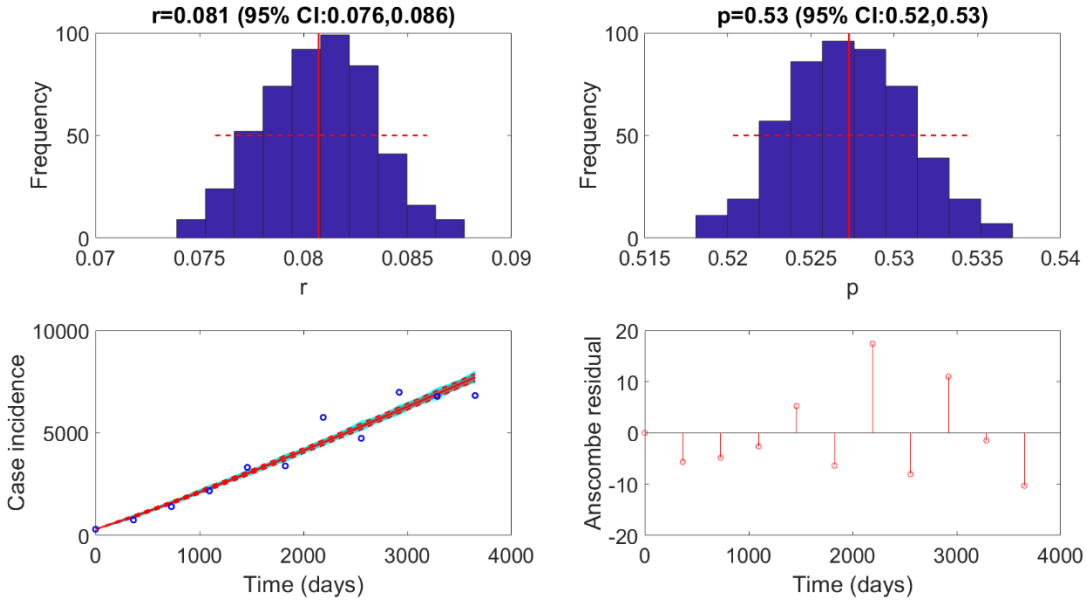
a-2 FMD (Uruguay, 2001) (10w) with MLE method



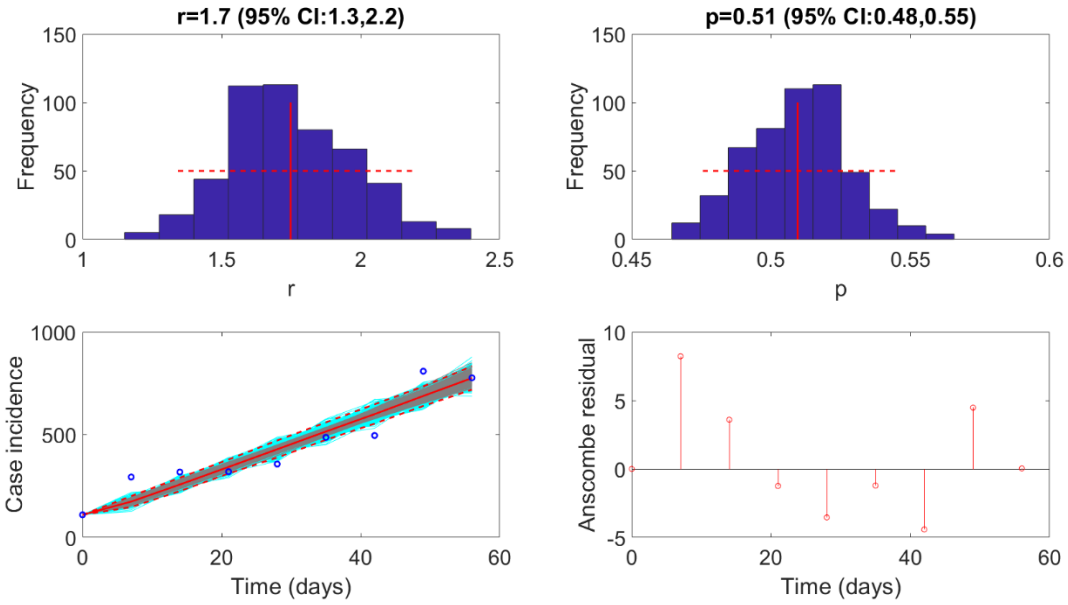
b-1 HIV-AIDS (Japan, 1985-2012) (11y) with LSQ method



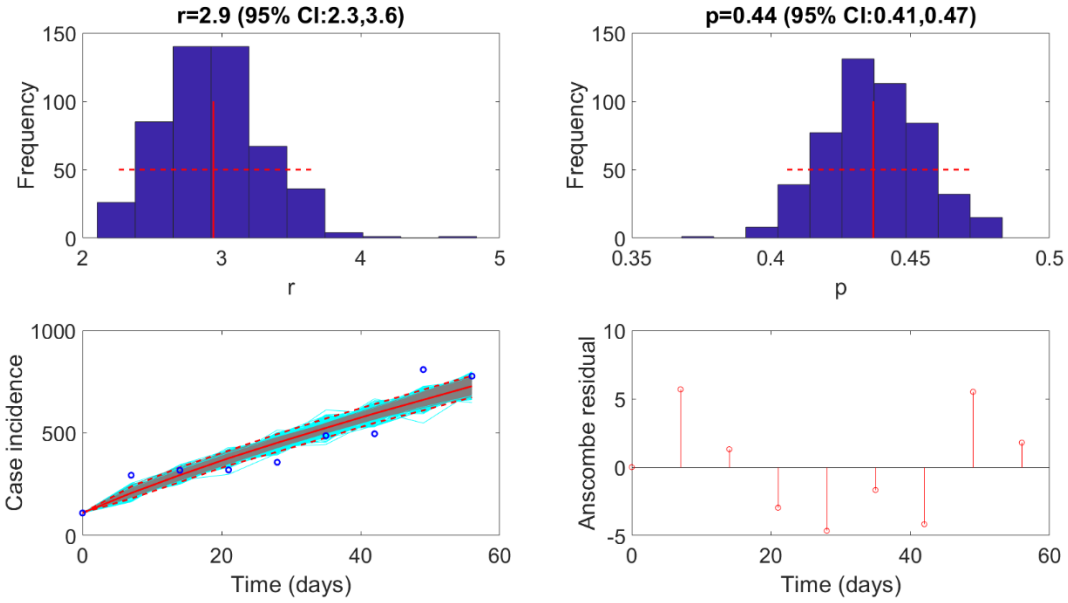
b-2 HIV-AIDS (Japan, 1985-2012) (11y) with MLE method



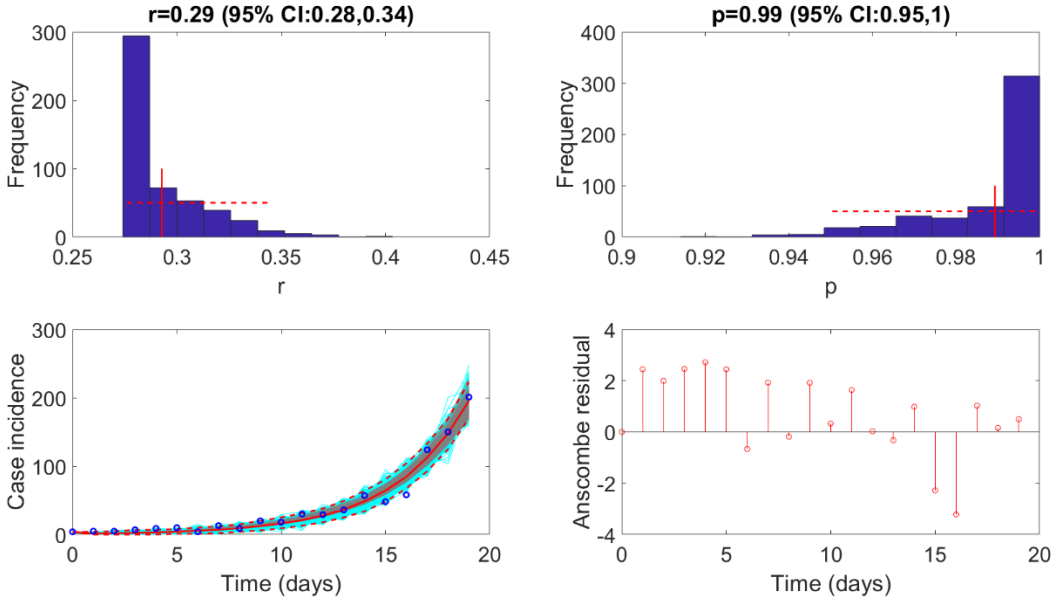
c-1 Measles (London, 1948) (9w) with LSQ method



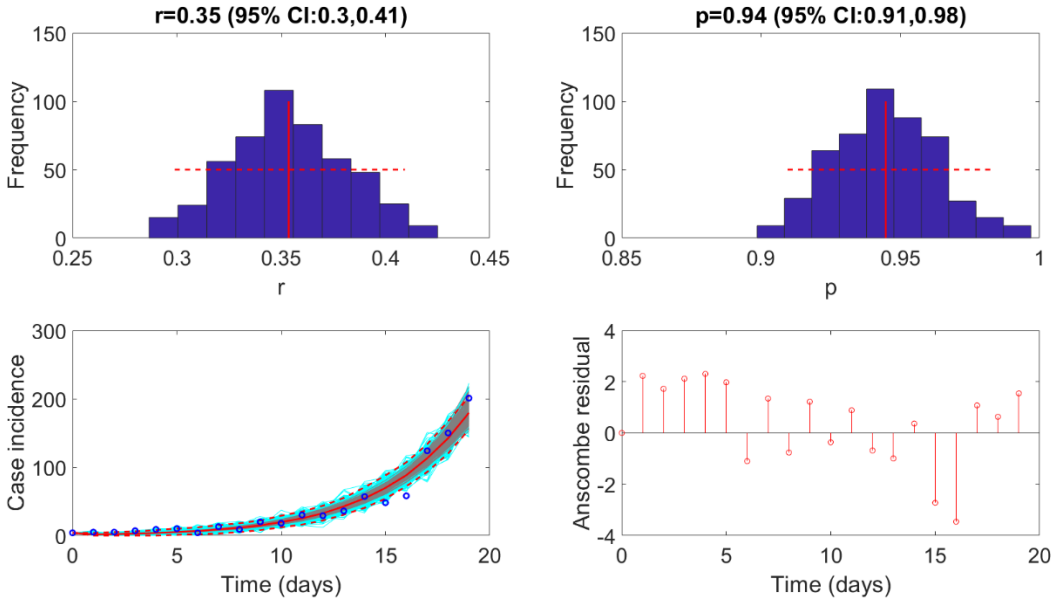
c-2Measles (London, 1948) (9w) with MLE method



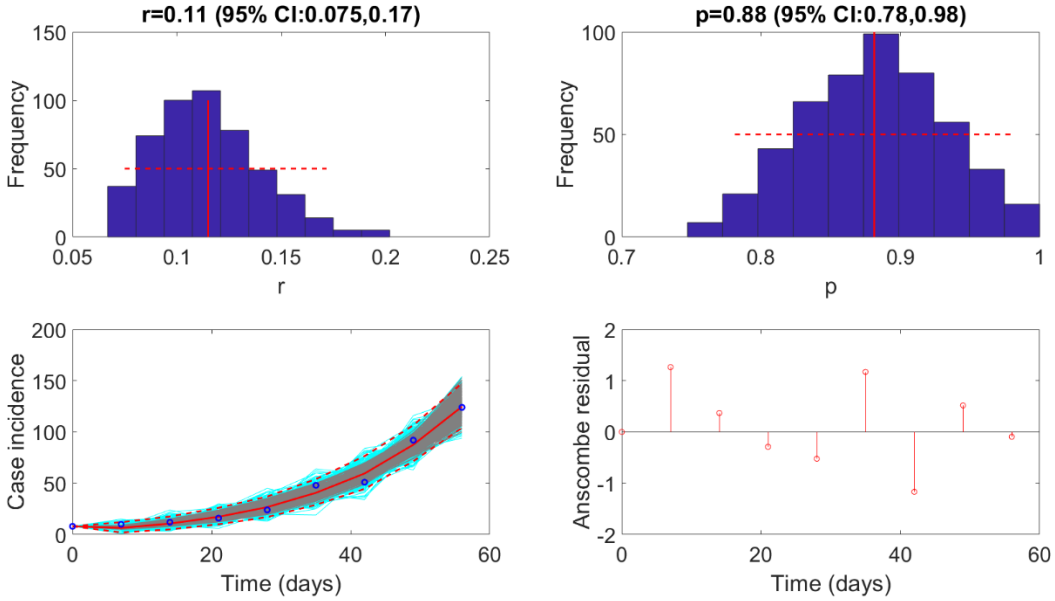
d-1 Pandemic influenza (San Francisco,1918) (20d) with LSQ method



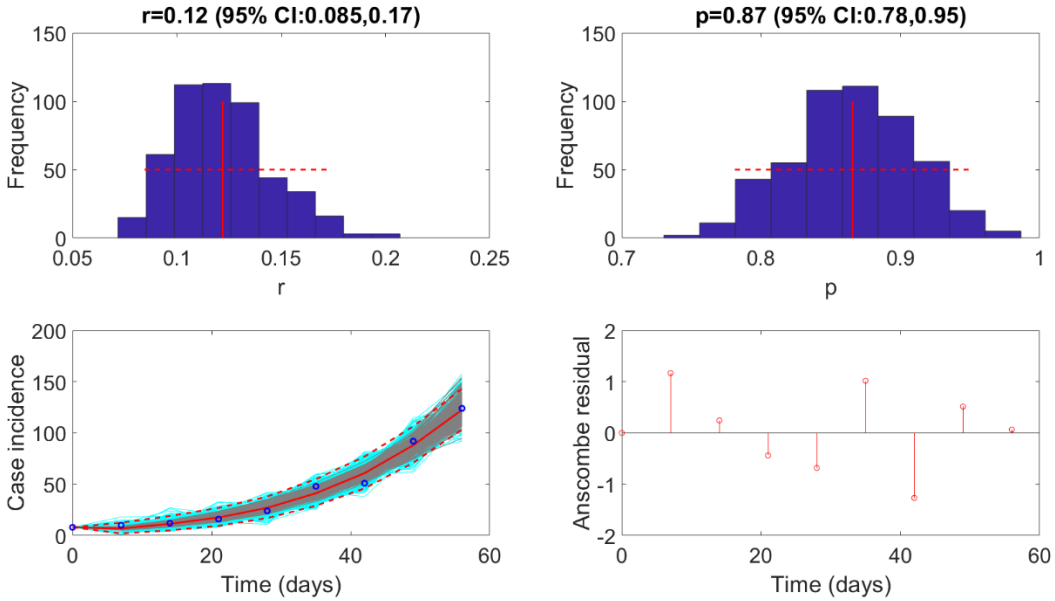
d-2 Pandemic influenza (San Francisco,1918) (20d) with MLE method



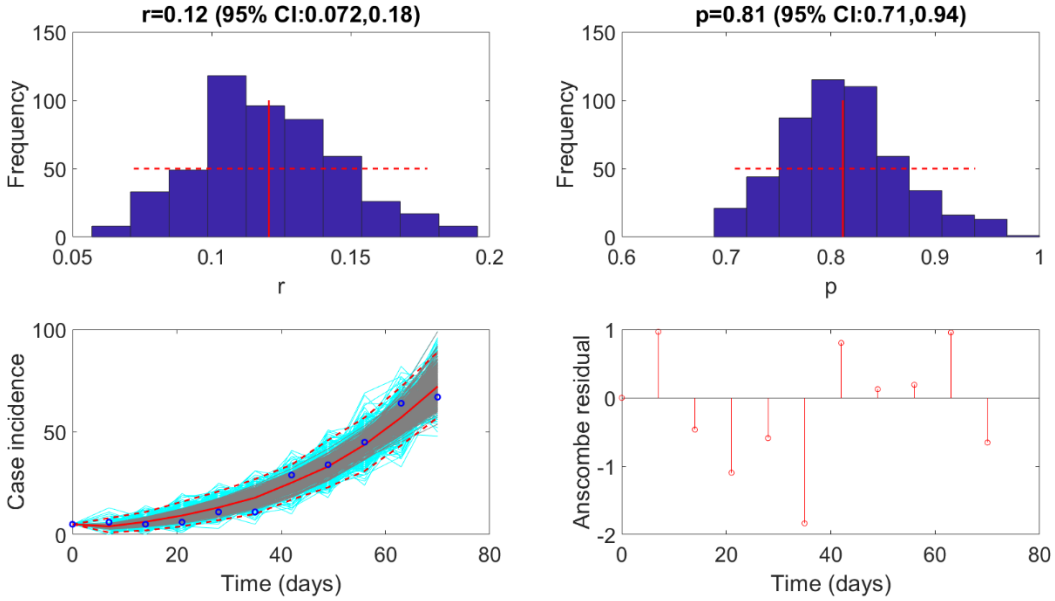
e-1 Plague (Bombay, 1905-06) (9w) with LSQ method



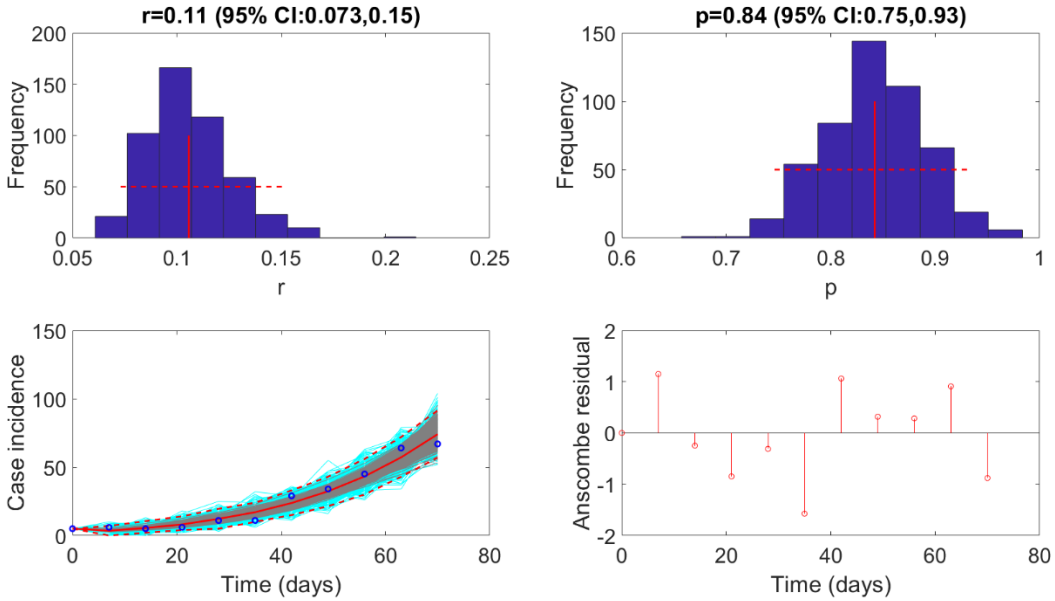
e-2 Plague (Bombay, 1905-06) (9w) with MLE method



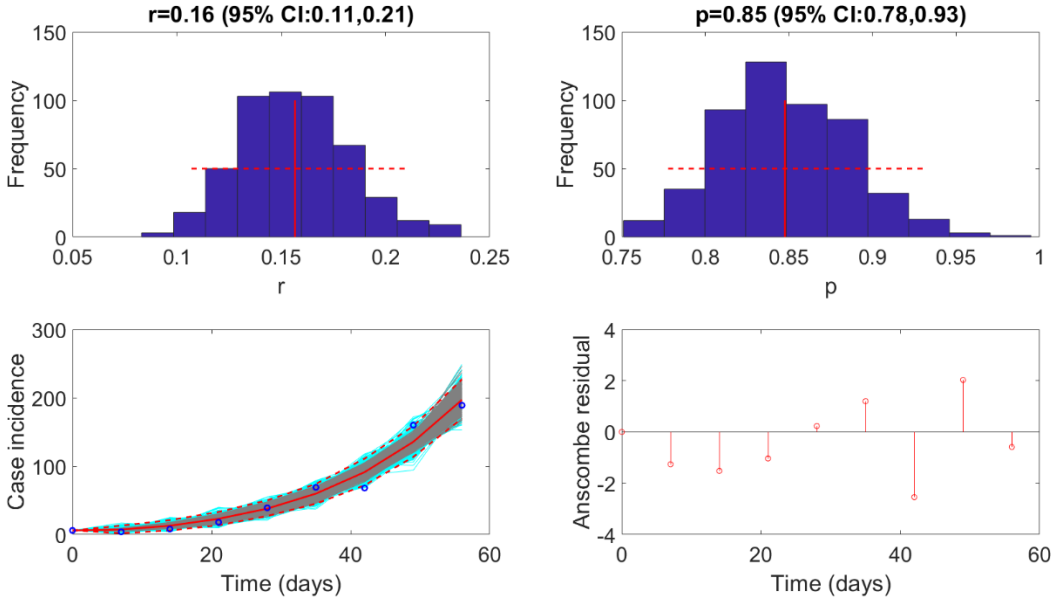
f-1 Plague (Madagascar-wave2, 2017) (11w) with LSQ method



f-2 Plague (Madagascar-wave2, 2017) (11w) with MLE method



g-1 Smallpox (Khulna, Bangladesh, 1972) (9w) with LSQ method



g-2 Smallpox (Khulna, Bangladesh, 1972) (9w) with MLE method